THE UNIVERSITY
of EDINBURGH

Understanding the relationship between IRF-1 and the transcriptional repressor ZNF350

Lucy Mallin

Doctor of Philosophy

The University of Edinburgh

October 2014

Declaration

I hereby declare that I am the author of this thesis. The work presented here was performed by myself and any contributions by others has been clearly indicated and acknowledged. I can confirm that no part of this thesis has been submitted for any other degree or professional qualification.

Lucy Mallin

October 2014

Acknowledgements

I would like to dedicate this thesis to my Grandad, who always inspired me to stick in and try my hardest. His motto “always read the question” has helped throughout my academic career and I’m sure will continue to do so throughout my life.

I would like to start off my thanking my supervisor Kathryn for her guidance and patience with me. I am very grateful to her for giving me the opportunity to study in her research group and for all the help and support provided along the way. Also a general big thank you to all the members of Kathryn and Ted’s groups for all their help in the lab as well as numerous entertaining tea break discussions. In particular, thanks to Vikram for helping me when I was starting out in the lab as well as Vivien, Yuh-Ping, Fiona, Fi and Kalai for help throughout my time there with experiments and insight. Also, thanks to my second supervisor Colin and members of the bioinformatics team including Graeme and GoGo, for helping me with all the data analysis. When I started out, I was a bit scared of computers but I gradually saw the light and developed a new love of bioinformatics. So much so that I have now undertaken a bioinformatics-based job so thank you for introducing me to the world of computational biology. Thank you to Richard, who I know was probably sick to death of stories of my cells dying, but has supported me endlessly and helped calm me down when things got a bit much. I can’t express how grateful I am that you allow me to drag you about the country for my jobs-I’ll pay you back some day. Thank you to my Mum and Dad for supporting me throughout my never-ending studies both financially and emotionally-I’m almost done I promise. Finally, to my awesome sister Sophie, thank you for the multitude of cup of tea visits and providing me with a nice distraction. They were really needed this past year and I’m really going to miss them and you over the next few years.

Abstract

Interferon regulatory factor-1 (IRF-1) is a transcription factor and tumour suppressor, involved in many diverse cellular processes including immune responses and growth regulation. An interesting feature of IRF-1 is that it can both activate and repress gene expression, possibly by acting with co-activator or co-repressor proteins. In a previous phage display assay, a homologous peptide to the known repressor protein, zinc finger 350 (ZNF350), was found to bind to the C-terminus of IRF-1. ZNF350, also known as ZBRK1 (Zinc finger and BRCA1-interacting protein with KRAB domain-1), is a member of the Krüppel-associated box (KRAB)-containing zinc finger (KZF) proteins, which is a group of the widely distributed transcriptional repression proteins in mammals. ZNF350 has previously been shown to repress the expression of a number of genes including *ANG1* and *GADD45A*, often in complex with other proteins. This study confirms the direct interaction between IRF-1 and ZNF350 and identifies key residues, including the LXXLL repression motif within the C-terminus of IRF-1, necessary for the binding interface. The two proteins have additionally been shown to interact within a cellular environment, shown by using techniques including immunoprecipitation and a proximity ligation assay. In addition, the ZNF350/IRF-1 complex formation appears to occur in the basal state of the cell, as opposed to in response to cellular stress such as viral infection or DNA damage. On the basis of ZNF350 being a negative regulator of transcription, a novel technique was developed to identify putative targets of both ZNF350 and IRF-1. This involved an initial bioinformatics screen using candidate IRF-1 binding site data obtained from CENTIPEDE, an algorithm that combines genome sequence information, with cell-specific experimental data to map bound TF binding sites. This allowed for the identification of novel target genes that contained the ZNF350 consensus binding site, GGGxxCAGxxxTTT, within close proximity to an IRF-1 consensus site, such as the immune response gene *IL-12A*. Lastly, a peptide phage display screen was combined with high-throughput sequencing to identify other potential binding partners of ZNF350 and perhaps help to understand the mechanism by which transcriptional repression is controlled by complex formation.

Abbreviations

aa	Amino acid
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
ATM	Ataxia-telangiectasia Mutated
Bax	Bcl2-associated protein X
bp	Base pair
BSA	Bovine serum albumin
CDK	Cyclin dependent kinase
ChIP	Chromatin immunoprecipitation
Co-IP	Co-immunoprecipitation
DBD	DNA binding domain
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ECL	Enhanced chemiluminesence
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
GST	Glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HRP	Horse radish peroxidase
IFN	Interferon

IPTG	Isopropyl-β-thio-galactoside
IP	Immunoprecipitation
IR	Ionising radiation
kb	Kilobase
kDa	Kilodalton
mAb	Monoclonal antibody
MEF	Mouse embryonic fibroblast
mRNA	Messenger RNA
NGS	Next generation sequencing
OD	Optical density
pAb	Polyclonal antibody
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pfu	Plaque-forming unit
pH	Potential hydrogen
RLU	Relative light unit
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
TAD	Transactivation domain
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
wt	Wild type
Zn	Zinc

Table of Contents

Figures

Tables

Chapter 1: Introduction

1.1 Transcription

1.1.1 Transcriptional Activation

Transcription is the first step in gene expression and involves the synthesis of RNA from a DNA template by the RNA polymerase enzyme. The transcribed RNA is then used to produce proteins via the process of translation. In order for transcription to begin, RNA polymerase II (Pol II) must associate with the DNA upstream of the open reading frame, where one of the strands then acts as a template for the synthesis of a complimentary RNA molecule [1-3]. One of the most important DNA sequences that signal the start of transcription is the TATA box, usually located 25 bp upstream of the transcription start site. In eukaryotic cells, unlike bacteria, general transcription factors (GTFs) are required to help position the Pol II correctly at the promoter. These GTFs consist of a set of interacting proteins designated as TFII and listed as TFIIA, TFIIB and so on. The first GTF to bind to the DNA is TFIID, which is a complex containing the TATA-box binding protein (TBP) and various TBP-associated factors (TAFs). The assembly of GTFs on the DNA template forms the pre-initiation complex (PIC). In cells, RNA Pol II may also be loosely associated with a variety of factors such as GTFs, regulatory proteins including Srb/Med proteins and chromatin modifying enzymes [1; 2]. This complex is known as the RNA Pol II holoenzyme and can be recruited to promoters without the prior assembly of GTFs [4].

Transcription also requires activator proteins which bind to specific DNA sequences and stimulate initiation and elongation. Often activators require the assistance of co-activator proteins, which do not have DNA binding activity but interact with the activators and help them carry out their function. They do this in a variety of ways, including interacting directly with components of the PIC to stimulate its formation or promoting the alteration of chromatin structure at the promoter. Activators and

co-activators can interact with chromatin-binding proteins, histone-modifying enzymes and ATP-dependent nucleosome-remodelling proteins and recruit them to the promoter. These proteins are associated with the RNA Pol II holoenzyme and can affect the interaction between histones and DNA, nucleosome association and the position of these nucleosomes in relation to the transcription factor binding site. By altering the chromatin structure, the DNA becomes more accessible for transcriptional machinery such as RNA Pol II and transcription factors [2; 5].

In order for transcription to begin, RNA Pol II must undergo a conformational change to release it from the promoter. A key step in this is the addition of phosphate groups to the tail of RNA Pol II (known as the C-terminal domain or CTD). The RNA polymerase can then uncouple from the general transcription factor complex, undergoing a series of conformational changes to allow for a tighter interaction with the DNA and recruitment of further proteins, allowing it to continue transcription for long distances without dissociating from the DNA template [5]. Once elongation has begun, most of the initiation complex proteins are released to enable them to begin another round of transcription with a new RNA Pol II molecule. During elongation, RNA polymerase II is associated with a series of proteins that make up the transcript elongation complex (TEC) and are important for a variety of steps of RNA processing such as pre-mRNA capping, splicing, 3'-end processing, surveillance and export [6]. Since DNA is tightly packed into nucleosomes, these have to be removed or remodelled in order for Pol II to progress with transcription. This includes nucleosome mobilization and histone depletion [7] and involves various factors such as the FACT (facilitates chromatin transcription) complex which has been shown to displace the histone H2A-H2B dimer from the nucleosome [8]. FACT has also demonstrated its involvement in nucleosome reassembly behind the transcribing RNA Pol II [9].

Transcription termination can occur through different pathways depending on the RNA 3'-end processing signals and termination factors at the end of the gene [10]. The newly synthesised mRNA is then exported from the nucleus for translation [5].

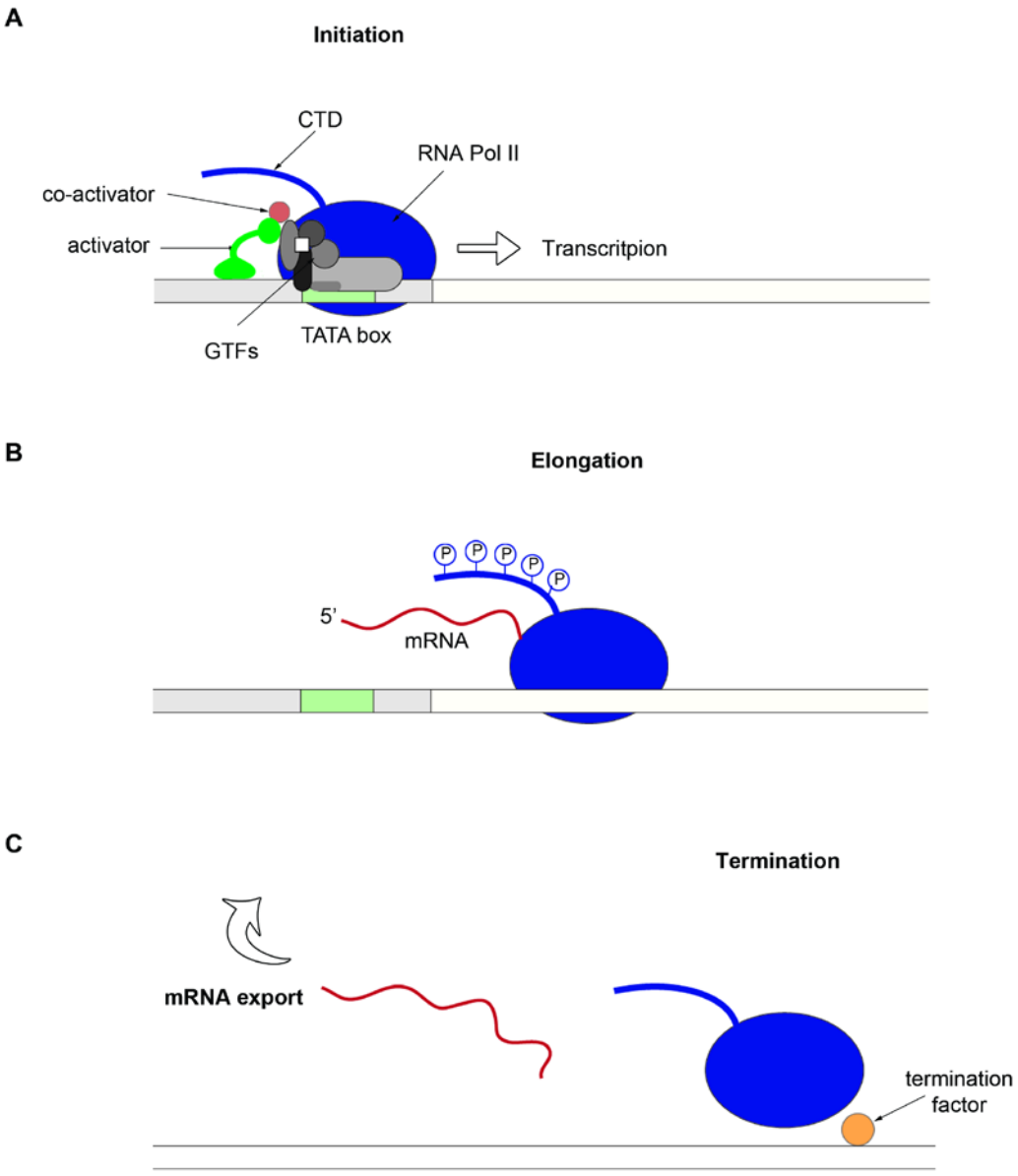


Figure 1-1: Overview of Transcription

(A) Binding of a transcription factor to its target promoter allows for the recruitment of general transcription factors (GTFs), co-activators, and RNA Polymerase II (Pol II) and subsequent complex formation at the promoter, leading to transcription initiation. (B) During elongation, RNA is synthesised by RNA Pol II. A key step in this is the addition of phosphate groups to the tail of RNA Pol II (CTD) which allows RNA Pol II to undergo a conformational change enhancing its association with DNA. (C) Termination of transcription occurs when RNAPol II associates with termination factors. Newly synthesised mRNA is exported out of the nucleus

1.1.2 Transcriptional Repression

Gene expression must be properly regulated for the control of cell growth, differentiation and homeostasis. This can be attributed to some transcription factors acting to repress expression and for most genes transcription initiation is the main point at which expression is regulated. Transcription repression falls into two categories: general (or global) repression and gene-specific repression. General repression occurs when a repressor protein or complex targets the components of PIC or RNA Pol II and modifies or sequesters them so they cannot function properly. Alternatively, gene-specific repression results from the activity of gene-specific repressors or co-repressors to inhibit transcription of a specific gene [11; 12].

1.1.2.1 Repressor Proteins

Gene-specific repressors can regulate transcription by either indirectly or directly binding to DNA close to the promoter (short-range) or a considerable distance up or downstream (long-range). 'Long-range' repression can occur when the repressor protein affects chromatin structure at or near the promoter [2]. Another suggested method of long-range repression involves the repressor protein being able to make contact with activators and/or components of the PIC by 'looping' of the chromatin [13] (Figure 1-2A). Conversely, 'short-range' repression takes place when the repressor acts locally at the promoter to affect transcription. Interestingly, a repressor protein may cause both long and short range repression depending on its binding site and mechanism of action [2; 11]

Repressor proteins act locally in a variety of ways to regulate specific genes. They can interrupt the activity or position of the activator or co-activator by inhibiting its binding to DNA (Figure 1-2B) or other transcriptional machinery or preventing the function of the DNA-bound activator/co-activator (Figure 1-2C). Alternatively, repressor proteins can regulate the turnover and levels of activator/co-activator and its localisation within the cell [2]. For example, MDM2 is an oncoprotein and

repressor that ubiquitinates the p53 transcription factor, triggering its export from the nucleus and subsequent degradation [14; 15].

Another route of repression is to target the transcription machinery, for example GTF interaction, interference of binding between core components of the PIC and modification of RNA Pol II [2]. The C-terminal domain (CTD) of RNA Pol II is a common target for repression since it undergoes certain modifications during transcription. During initiation, the CTD is glycosylated and de-phosporylated whereas during elongation it is de-glycosylated and phosphorylated. By altering the progression of these modifications, the repressor protein can repress transcription [16].

1.1.2.2 Chromatin Remodelling

A third method of repression is via the remodelling and compaction of chromatin. Chromatin is a complex of DNA and protein that consists of DNA wrapped around histones H2A, H2B, H3 and H4 forming nucleosomes, which themselves form higher order structures. These higher structures of chromatin appear in two forms: euchromatin and heterochromatin. Euchromatin, also known as decompacted chromatin, contains genes that are usually expressed in cells whereas heterochromatin tends to be highly compacted and therefore contains a small number of active genes. The switching of chromatin between the active and repressive states can result from the modification of the N-terminal tails of histones, particularly H3 and H4. Histones can be modified at several sites within their tail by acetylation [17], methylation [18] or phosphorylation [19]. Other modifications such as ubiquitination [20; 21] and sumoylation [22] can also occur but are less well understood.

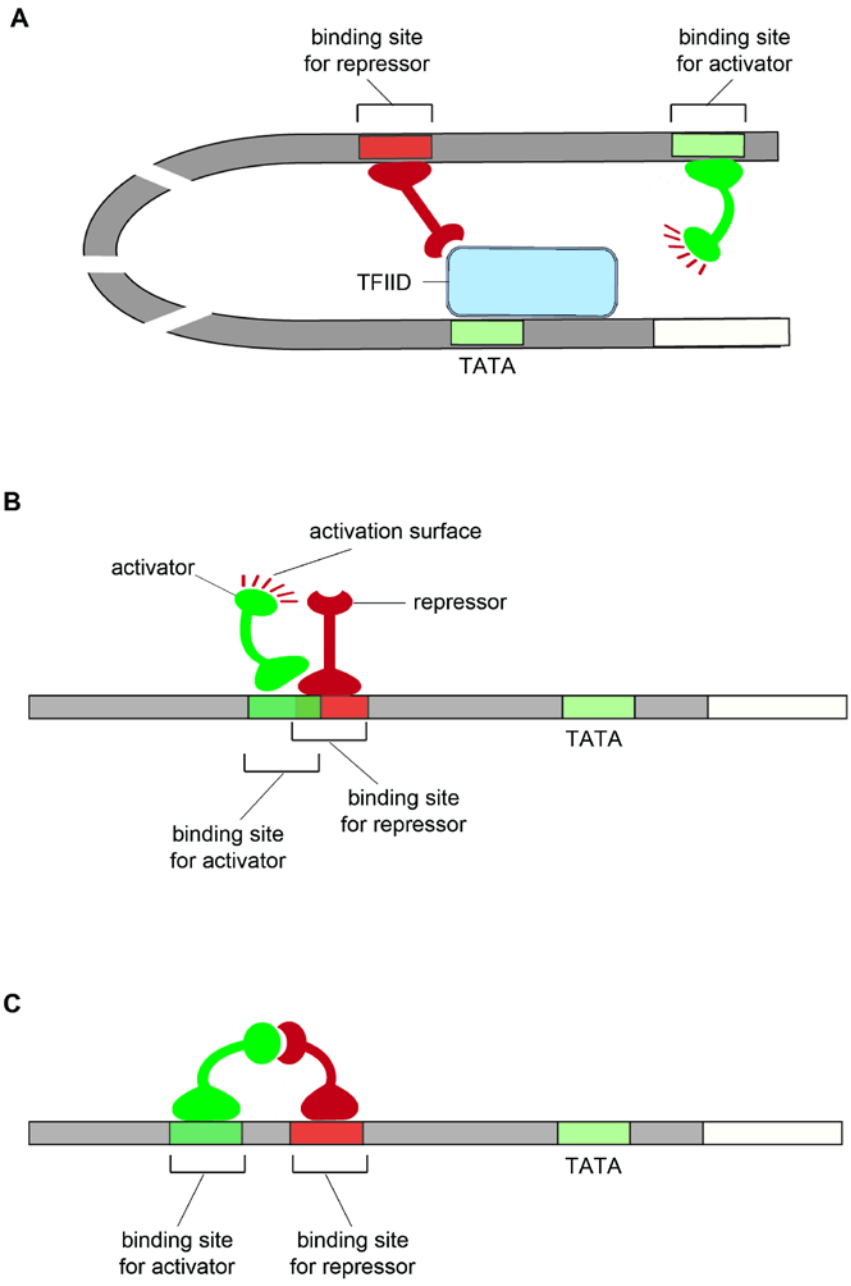


Figure 1-2: Activity of gene repressor proteins (adapted from Alberts et al. 2002 [5])
(A) DNA "looping". The repressor acts to prevent the assembly of the complex of general transcription factors by binding to early components of the complex such as TFIID, thereby preventing further assembly. (B) Gene activator proteins and gene repressor proteins compete for binding to the same regulatory DNA sequence. (C) Both proteins can bind to the DNA but the repressor prevents the function of the activator by binding to its activation domain.

Acetylation is a reversible process carried out by enzymes known as histone acetyl transferases (HATs) which acetylate conserved lysine residues on histone N-terminal tails. This process removes the positive charge on the histones, which decreases the interaction with the negatively charged phosphate groups of DNA. As a result, the chromatin structure becomes more relaxed, enabling an increase in gene transcription. This process can be reversed by histone deacetylases (HDACs) which remove the acetyl groups from histones, increasing their interaction with DNA and ultimately condensing chromatin rendering it mostly inactive [23].

Another post-translational feature found on inactive, condensed chromatin is histone methylation. This process is carried out by histone methyltransferases (HMT) and involves the addition of methyl groups onto lysine or arginine residues of the H3 or H4 histones [23]. DNA methylation is also linked to histone modification and the repression of transcription. DNA is methylated by DNA methyltransferases (DMT), with most methylated sites located in regions known as CpG islands [2]. CpG islands contain a high frequency of CpG sites, which are cysteine nucleotides positioned next to guanine nucleotides. The cysteine within this dinucleotide is a common target for methylation. CpG methylation can bring about the repression of transcription by directly blocking the binding of transcriptional activators or indirectly by altering chromatin structure and/or recruiting methyl-CpG-binding proteins. These proteins have been shown to compete with transcription factors for binding to CpG-methylated DNA as well as recruiting HDACs to bring about changes in chromatin conformation [2]. A summary of some methods of chromatin remodelling can be seen in Figure 1-3.

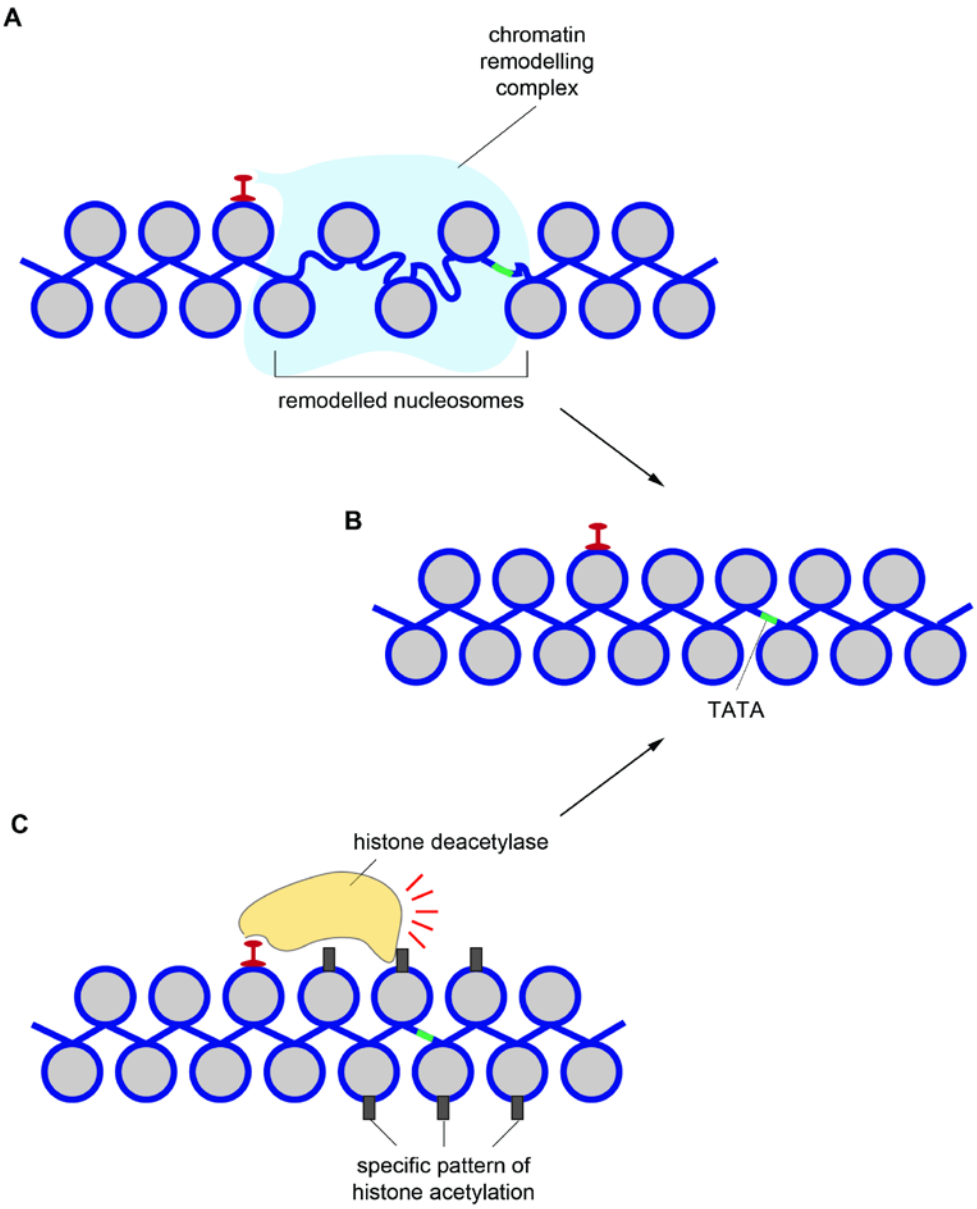


Figure 1-3: Chromatin remodelling

(A) The repressor recruits a chromatin remodelling complex, which returns the region to a condensed nucleosomal state (B), thereby preventing transcription. (C) The repressor attracts a histone deacetylase to the promoter, which removes actyl groups from histones and causing the chromatin to become condensed. Figure adapted from Alberts et al. 2002 [5].

1.2 KRAB-containing zinc finger proteins

The wide diversity of repression mechanisms described above highlights the probability that proteins involved in transcriptional repression may be structurally diverse. Repressor proteins that have a common mechanism will likely have similar repression motifs, an example of this being the Krüppel-associaetd box (KRAB) - containing zinc finger repressor proteins.

Zinc finger proteins are some of the most abundant proteins in the eukaryotic genome and are characterized by a zinc-binding motif of conserved cysteine and histidine residues. A finger protein typically consists of a series of zinc fingers made from a loop of ~23 amino acids protruding from the zinc-binding site [23]. Zinc finger proteins are structurally diverse and have a wide variety of functions including DNA- or RNA-binding, protein-protein interactions, membrane association, regulation of apoptosis, protein folding and assembly and transcriptional regulation [24].

The largest family of zinc finger transcription factors comprises those containing the Krüppel-associated box (KRAB) domain. This domain gets its name from the gap gene, Krüppel, which was originally identified in Drosophila melanogaster as encoding a zinc finger transcription factor [25]. KRAB domains are involved in transcriptional repression and it has been estimated that approximately one third of all zinc finger proteins present in the human genome are those containing the KRAB-domain, making them the largest family of transcriptional regulators in mammals [26]. To date, 742 structurally different KRAB- containing zinc finger proteins generated from 423 genes, have been identified. Many of these genes are found in clusters with most of them residing in 25 major clusters throughout the genome. The most well documented cluster is on chromosome 19 where over 200 KRAB-containing zinc finger gene loci are found at 11 different sites [27]. Other clusters have been found in centromeric and telomeric regions of other chromosome and non-clustered genes appear to be scattered throughout the genome with half on sex chromosomes and half on autosomes [28]. Interestingly, KRAB-containing zinc finger proteins only appear to be present in the genomes of tetrapod vertebrates.

Zinc finger proteins in *Drosophila*, fish, plants yeast and fungi lack the KRAB domain whilst in human, mouse, chicken frog and rat genomes, it is present [29].

1.2.1 Structural Features

Members of the KRAB-containing zinc finger protein family are made up of two common structural domains; C_2H_2 zinc fingers and a KRAB domain. Some members of the family may also possess a SCAN domain or a BTB motif.

1.2.1.1 Zinc Fingers

The Cys_2His_2 (C_2H_2) zinc fingers found in KRAB-containing proteins, as well as other zinc finger proteins, are the most well characterized of the zinc finger family. They adopt a $\beta\beta\alpha$ fold and are defined by the presence of the consensus sequence Z-X-Cys-X_{2-4}-Cys-X_3-Z-X_5-Z-X_2-His-$X_{3,4}$-His, where X represents any amino acid and Z represents a hydrophobic residue [28]. Many of the proteins containing this type of zinc finger are transcription factors that recognise and bind to specific DNA sequences [24]. In these proteins, zinc finger domains characteristically occur as tandem repeats at the C-terminus, with amino acids on the surface of the α-helix of each finger recognising a three-nucleotide sequence along the major groove of the DNA helix. It has been seen in some cases that not all zinc finger domains in the protein are involved in DNA binding, suggesting they may have additional roles. The linker region that separates neighbouring C_2H_2 zinc fingers is a really important structural element to help control the spacing of the fingers along the DNA binding interface. This region usually takes the form of TGEKPYX (with x representing any amino acid) and is required for high affinity DNA binding [24; 30]. KRAB-containing proteins often contain a high number of zinc fingers, on average 12, with proteins being known to contain 30 or more [27; 31].

1.2.1.2 KRAB domain

The KRAB domain is located near the N-terminus, spanning 50-75 amino acids and consists of one or both of the KRAB A and B boxes. The A box plays a key role in repression by binding to co-repressors whereas the B box enhances this A-box-mediated repression. These boxes are encoded by two individual exons separated by various sized introns. This exon-intron composition allows for alternative splicing resulting in different products [28]. The KRAB-containing zinc finger proteins have been divided into three subfamilies depending on this alternative splicing and consequent structure of the N-terminal domain. One subfamily consists of proteins containing an A box alone (KRAB A), the other with a combination of A and B boxes (KRAB A + B) and the third with an A box combined with a different B box, sometimes known as the b box (KRAB A + b) [29].

1.2.1.3 Other domains

The N-terminus of some KRAB-containing zinc finger proteins may also contain a SCAN or BTB domain. The SCAN domain is named after the founding members of the family **S**RE-ZBP, **C**Tfin51, **A**W-1 and **N**umber 18 cDNA but is also known as LeR due to its leucine-rich primary structure. This domain is a highly conserved region containing at least 87 residues and mediates protein-protein interactions, often with other SCAN zinc finger proteins [32]. The BTB (**B**road-complex **T**ramtrack **B**ric-a-brac) domain is another conserved motif involved in dimerization and has been shown to be important for the proper folding of the zinc finger protein [33].

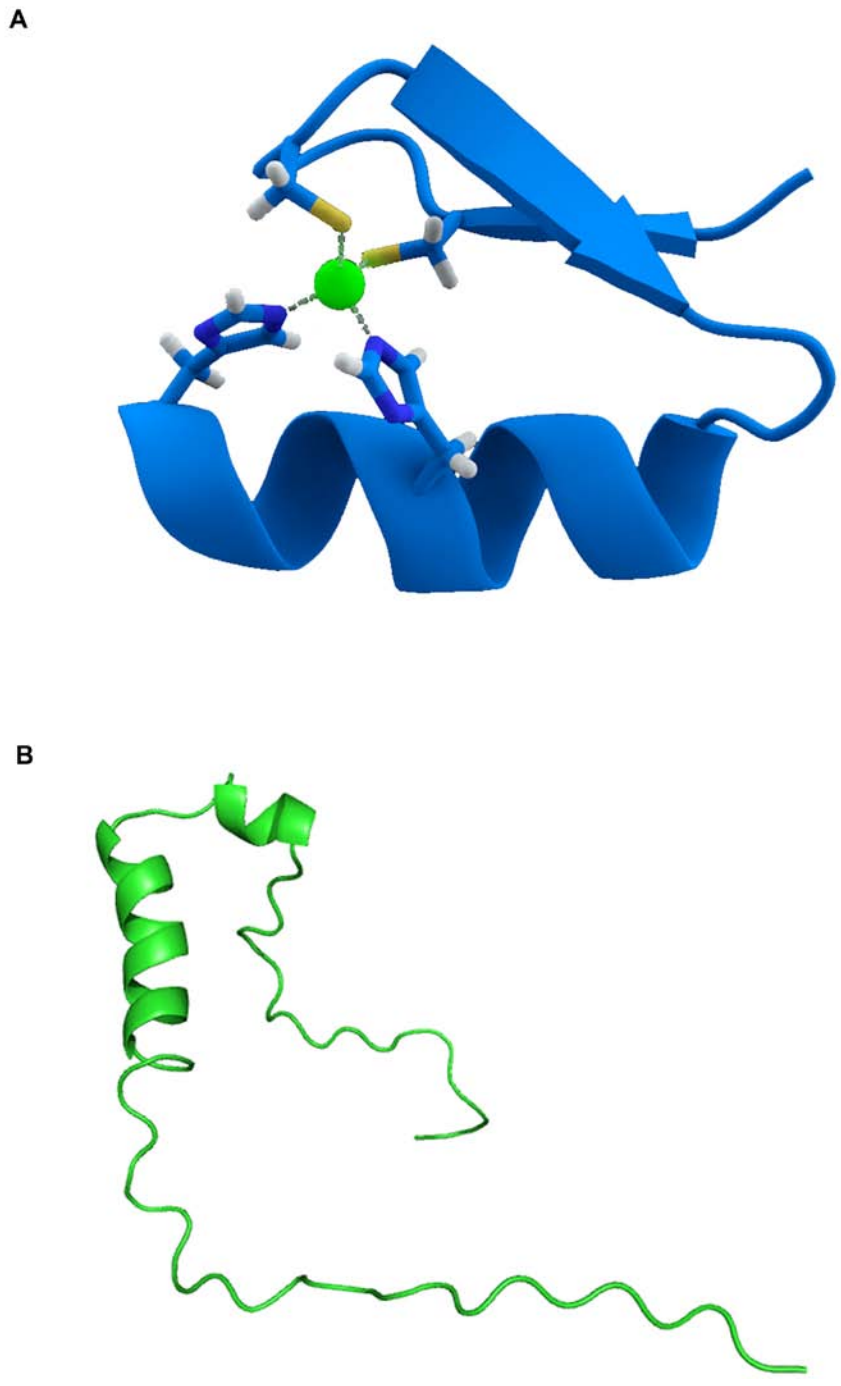


Figure 1-4: Structural features of KRAB-ZFPs

(A) Schematic representing the Cys2His2 zinc finger motif, consisting of a α helix and an anti-parallel β sheet. The zinc ion (green) is coordinated by two histidine residues and two cysteine residues. Based on the X-ray structure of PDB 1A1L and generated by Thomas Splettstoesser [34] (B) Representation of the Kruppel Associated Box (KRAB) domain, produced by Sean (2007) [35].

1.2.2 Function and mechanism

KRAB-containing zinc finger proteins repress gene-specific transcription by interacting with chromatin-remodelling factors. The zinc fingers bind to the corresponding DNA sequence while the KRAB domain interacts with a co-repressor protein KAP1 (KRAB-associated protein-1). KAP1, also known as TRIM28, has an N-terminal tripartite motif (TRIM) which is composed of a RING, two B-box zinc fingers and a coiled-coil (RBCC) [36]. This motif binds the KRAB domain of zinc finger proteins as a homotrimer. The central region of KAP1 contains a hydrophobic pentapeptide, involved in the binding of heterochromatin protein-1 (HP-1) [37]. The C-terminus of KAP1 is the location of the tandem PHD and bromodomain (known as the PB domain) which act as a scaffold to recruit histone deacetylases, chromatin remodelling complexes (such as the NuRD complex) and histone lysine N-methyltransferases (such as SETDB1) to the promoters of target genes. This allows for the ATP-dependent modification and condensing of chromatin and ultimately gene repression [38-40]. A summary of KAP1-mediated transcriptional repression can be seen in Figure 1-5.

The functions for most of the KRAB-containing zinc finger family are not well established, however some known functions studied include repression of RNA Pol I, II and III promoters, binding and splicing of RNA, control of nucleolus function, cell differentiation, proliferation and apoptosis [28].

1.3 ZNF350

ZNF350, also known as ZBRK1 (Zinc finger and BRCA1-interacting protein with KRAB domain-1) is a member of the KRAB-containing zinc finger protein family and was initially identified by a yeast two-hybrid screen as a protein associated with BRCA1 [41]. A full length cDNA clone was isolated that encoded a protein of 532

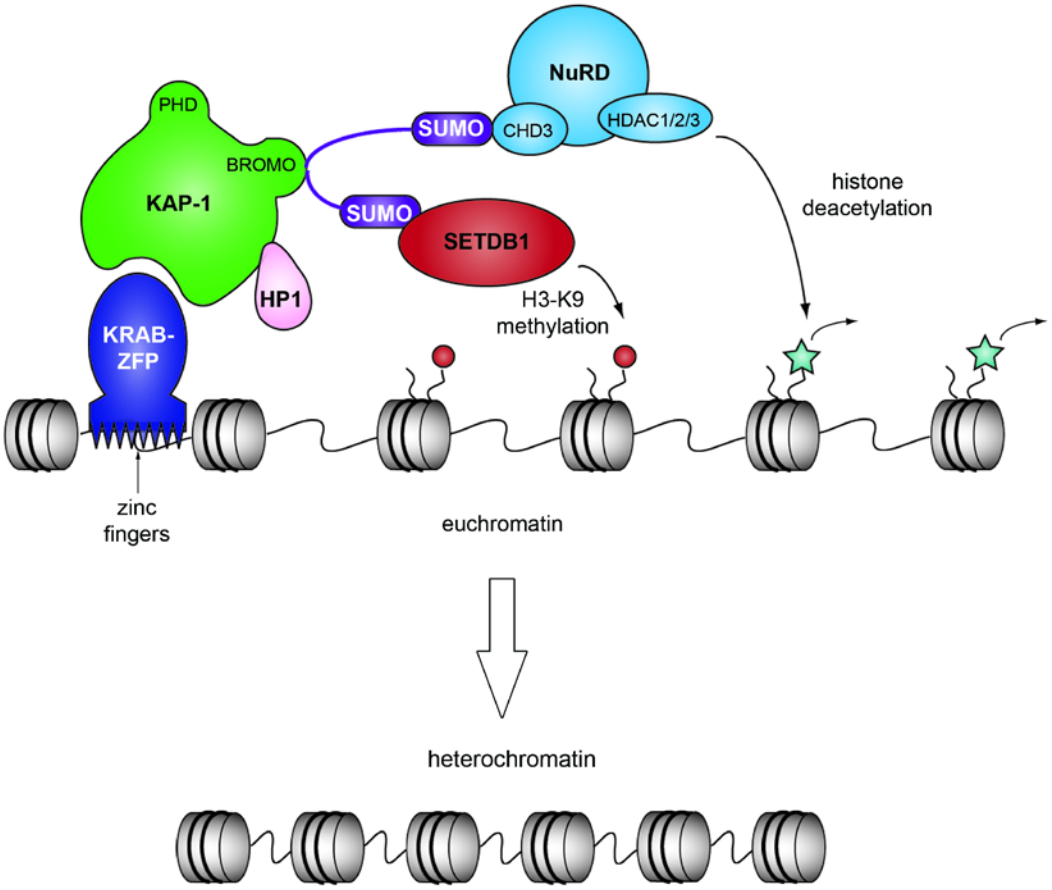


Figure 1-5: Mechanistic action of KAP1as a co-repressor

KRAB-ZFP binds to target DNA via zinc fingers and recruits the co-repressor, KAP1. The gene is still being actively expressed and is associated with acetylated nucleosomes. The KAP1 PHD finger functions as an E3 ligase, physically interacting with and directing the E2 enzyme to SUMOylate the KAP1 bromodomain. The SUMO-bromodomain recruits the histone lysine N-methyltransferase SETDB1, which methylates lysine 9 of histone H3. The SUMO-bromodomain also recruits the NuRD complex, which deactylates nucleosomes associated with the KRAB-ZFP target gene. Consequently, chromatin becomes condensed, forming heterochromatin and gene expression is repressed. Figure adapted from [42].

amino acids and further analysis revealed an N-terminal KRAB A + B box domain and eight C_2H_2 zinc fingers in the central region [43; 44]. Further analysis of the eight zinc fingers of ZNF350 has revealed dual roles in DNA binding and protein interaction. Zinc fingers 1-4 are essential for DNA binding activity and recognise a compositionally flexible 15 bp DNA sequence GGGxxxCAGxxxTTT (with x being any nucleotide) [43]. Zinc fingers 5-8 appear to modulate the stability of this DNA association but are also importantly involved in associating with co-repressors including BRCA1. A remarkable feature of ZNF350 that sets it apart from other KRAB-ZFPs is its possession of a C-terminal repression domain (CTRD) in addition to the N-terminal KRAB domain. Whilst the KRAB repression domain can interact with KAP-1 and exhibits broad promoter specificity, the CTRD appears to be BRCA1-, histone deacetylase- and promoter-specific [45]. The CTRD of ZNF350 has also been shown to undergo tetrameric oligomerization which may facilitate ZNF350-directed transcriptional repression. Generally, oligomerization allows for more diversity among sequence-specific DNA-binding transcription factors by expanding the number of response elements to which they can bind or alternatively, increasing the number and type of effective regulatory domains at the promoter [46].

1.3.1 Localisation

ZNF350 has been shown to localize within the nucleus along with BRCA1, but is completely excluded from the nucleolus [43]. Another study has shown that ZNF350 associates with the nuclear matrix and interestingly, when a ZNF350 mutant minus the KRAB domain is transfected into cells, about half of the protein no longer associates with the nuclear matrix but instead appears to associate with chromatin [44].

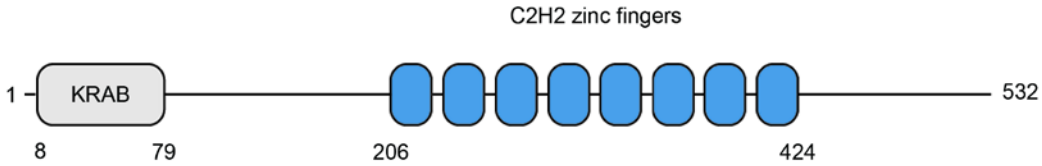


Figure 1-6: Functional domains of ZNF350

ZNF350 has an N-terminal KRAB domain (residues 8-79), eight zinc fingers (residues 206-424) and a C-terminal repression domain (CTRD).

1.3.2 Mechanism of Action

1.3.2.1 Repression

1.3.2.1.1 GADD45A

The breast and ovarian cancer susceptibility gene *BRCA1* is a tumour suppressor gene implicated in familial and sporadic breast cancer [47]. BRCA1 participates in several different complexes to coordinate a diverse range of cellular functions including transcriptional regulation, DNA repair, cell cycle control, apoptosis and centrosome duplication to maintain genomic stability [48]. In many of these complexes, BRCA1 functions as an E3 ubiquitin ligase by associating with another protein BARD, and directs the ubiquitination of specific targets within each complex [49]. BRCA1 has been implicated in the regulation of DNA damage-inducible genes that function to arrest the cell cycle. The *GADD45A* gene is transcriptionally induced by the over expression of BRCA1 and has been implicated in regulation of growth arrest processes such as activation of DNA damage-induced G2/M checkpoints [50]. Located within intron 3 of this gene is a functionally important p53 response element and a ZNF350 consensus binding motif downstream. ZNF350 specifically binds this consensus site in complex with BRCA1 and represses the expression of the *GADD45A* gene. The repressor activity of ZNF350 is dependent on BRCA1 and this repression is relieved by BRCA1 over-expression [43; 45]. Additionally, ZNF350 repressor activity is dependent on HDAC activity and it has been proposed that since the C-terminus of BRCA1 can interact with HDACs, ZNF350 recruits BRCA1 to genes as part of a repression complex that minimally includes an associated HDAC activity [45]. In a separate study, ZNF350 was found to bind to the same intronic site on the *GADD45A* gene in complex with KAP1 and another coregulator, ligand-dependent co-repressor (LCoR). This binding resulted in the repression of *GADD45A* expression in malignant and non-malignant breast epithelial cells and deletion of any of the complex causes increased expression of *GADD45A* and apoptotic cell death [51]. The observation that ZNF350 can interact in two separate repressor complexes on the same gene is quite interesting and further

studies will be required to understand under what circumstances each complex if formed.

1.3.2.1.2 ANG1

The *ANG1* gene encodes the protein angiopoietin-1 which has important roles in vascular development and angiogenesis. Interestingly, there are five potential ZNF350 consensus binding motifs in the *ANG1* gene, however only one, located in the promoter, actually confers ZNF350 binding. Through this site, ZNF350 binds in complex with BRCA1 and the co-repressor CtIP and represses *ANG1* expression. All three components of the complex are required for DNA binding and if just one is depleted the entire complex, including ZNF350, dissociates, which implies that the stability of ZNF350 binding is dependent on the integrity of the entire complex. When a point mutation is introduced into the ZNF350 binding site on the *ANG1* gene, transcription is upregulated, suggesting activation of the region possibly by the reorganization of local chromatin to allow for the binding of transcriptional activators. Studies in mice have shown that *Brca1*-deficient mammary tumours express high levels of *Ang1*. Interestingly, the promoter of the mouse gene contains an analogous ZNF350 recognition site, which may suggest a similar mode of regulation to that found in humans [52].

1.3.2.1.3 HMGA2

The *HMGA2* gene encodes the high mobility group AT-hook 2 (HMGA2), a DNA architectural protein that has important roles in tumorigenesis and is overexpressd in many different types of tumours. The HMGA family of proteins are responsible for the formation of multiprotein complexes on promoter/ enhancer regions and remodelling of chromatin resulting in the regulation of gene expression [53]. HMGA2 in particular is critical for various cellular processes such as gene transcription and promotion of metastatic progression and the repression of *HMGA2* is thought to prevent mammary tumourigenesis [53; 54]. The *HMGA2* promoter

contains a consensus ZNF350 binding motif upstream of exon 1 which binds ZNF350 in complex with BRCA1 and CtIP. Similar to the repression mode of *ANG1,* if there is a defect in ZNF350/BRCA1/CtIP repressor complex formation or depletion of any the three proteins, repression of *HMGA2* is relieved and there is a subsequent induction of gene expression. This leads to increased proliferation, anchorage-independent growth in soft agar and impairment of mammary acini formation highlighting a direct link between the oncogenic activity of HMGA2 and the ZNF350/BRCA1/CtIP complex [54].

1.3.2.1.4 p21

Cell cycle regulator $p21^{waf1/cip1}$ contains three ZNF350 consensus binding motifs at its promoter to which ZNF350 binds in complex with KAP1. Full KAP1 repressor activity usually requires sumoylation, however its association with the KRAB-domain of ZNF350 and binding to the *p21* promoter can occur regardless of its sumoylation state. The binding of ZNF350/KAP1 complex leads to the repressive modification of histone tails, presumably by the recruitment of SETDB1 to KAP1and *p21* expression is repressed [55].

1.3.2.1.5 HIV-1 LTR

HIV-1 is able to insert its genetic material into host genomes through the activity of its long terminal repeat (LTR). Recently, ZNF350 has been shown to bind to a consensus motif in the HIV-1 LTR and directly suppress its transcriptional activity. Additionally, the study showed that this repression requires KAP1 and HDAC2 activity [56]. Previous reports had shown that KAP1 can suppress endogenous retroviruses by recruiting SETDB1, a H3K9 methyltransferase and HP1 [57; 58], however they were shown to have no significant effect on the repressor activity of ZNF350 on the HIV-1 LTR. Furthermore, the depletion of endogenous ZNF350 leads to the activation of HIV-1 LTR but in fact only a 25% reduction in the level of KAP1 at the LTR suggesting KAP1 may bind itself [56]. This is a really interesting

finding highlighting the possibility that the stimulation of KRAB-zinc finger proteins might assist in the development of antiviral therapies.

1.3.2.1.6 MMP9

Matrix metallopeptidase-9 (MMP9) is a metastatic protein involved in the breakdown of the extracellular matrix. ZNF350 can bind to a putative ZNF350 site in the promoter of the *MMP9* gene to and repress its activity and a loss in ZNF350 expression has been inversely correlated with the elevated expression of MMP9 in cervical cancer specimens. Interestingly, ZNF350 expression has been found to be significantly lower in cervical cancer cells and in 75% of cervical cancer specimens compared to normal tissue. Additionally, a reduction of ZNF350 expression has been seen in hepatocellular carcinomas and colorectal cancers. These findings suggest that ZNF350 may acts as a tumour suppressor with important roles in metastasis and tumourigenesis [59]. It has yet to be elucidated if ZNF350 acts alone in this instance of gene repression, or with a complex of proteins such as BRCA1 or KAP1.

1.3.2.1.7 FGF2

A novel ZNF350 binding site has recently been located in the fibroblast growth factor-2 (*FGF2*) gene between exons 1 and 2. In breast epithelial and adenocarcinoma cells, ZNF350 binds in complex with KAP1 and LCoR to an almost identical consensus motif to that found in the *GADD45A* gene and represses the expression of *FGF2.* When this complex is bound to *FGF2,* SETDB1 is also recruited and enrichment of the repressive mark H3K9Me3 can be observed. When the ZNF350/KAP1/LCoR complex is disrupted or deleted, *FGF2* expression is enhanced and apoptotic cell death occurs [51].

1.3.2.1.8 KAP1

The ability of ZNF350 to suppress cell growth appears to require interactions with both the N-terminal KRAB-domain and the C-terminal repression domain. Interestingly, exogenous expression of KAP1 only had a small effect on cell proliferation which may imply that other proteins could interact with the N-terminus of ZNF350. The N-terminal KRAB-domain of ZNF350 is also essential for inhibition of cell migration and invasion. KAP1 can promote cancer cell invasion and metastasis without affecting cell proliferation and an inverse correlation of ZNF350 and KAP1 expression levels has been seen in samples of aggressive cervical cancer with ZNF350 expression levels being significantly decreased while KAP1 levels were significantly increased. This led to the discovery that ZNF350 can directly bind to the promoter of KAP1 and repress its transcription thereby attenuating KAP1-induced invasion and metastasis [60].

1.3.2.2 Activation

1.3.2.2.1 SCA2

The *SCA2* gene product Ataxin-2 (ATXN2), is a protein thought to be involved in the polyglutamine disorder, spinocerebellar ataxia type 2 (SCA2), and has been linked to endocytotic processes and RNA metabolism. A yeast two-hybrid screen using ataxin-2 as a bait protein isolated ZNF350 as a binding partner. Further studies showed that the two proteins interact via ZNF350's C-terminal repression domain and they bind in complex to a ZNF350 binding motif in the *SCA2* promoter itself. What is most remarkable about this interaction however is the observation that the ZNF350/ataxin-2 complex actually activates the transcription of the *SCA2* gene, suggesting ataxin-2 acts as a co-activator. ZNF350 over expression was found to increase ataxin-2 protein levels whereas depletion or interruption of binding resulted in reduced ataxin-2 levels. This correlation was also seen in colon tumour tissues where a decrease of both ZNF350 and ataxin-2 levels were observed [61]. This is the first and so far only instance where ZNF350 has been shown to activate gene

expression, however other studies have revealed some KRAB-containing zinc finger proteins can activate gene transcription. For example, ZNF480 has been shown to positively regulate MAPK-mediated signalling pathways, resulting in the activation of AP-1 and SRE [62].

1.3.3 ZNF350 mRNA expression

The activity of ZNF350 as a repressor protein appears to be related to tumourigenesis as well as invasion and metastasis with the levels of expression of *ZNF350* being extremely important for proper regulation of cellular processes. *ZNF350* mRNA expression is found to be altered in 75% of primary breast carcinomas with 46% of these showing clear under-expression [63]. When *ZNF350* mRNA expression was compared to BRCA1 expression, there did not appear to be any correlation. There was also no significant association between altered *ZNF350* mRNA expression and tumour phenotype, suggesting the biological effect of altered *ZNF350* expression to be overcome by other molecular changes [63]. A similar under-expression of *ZNF350* was found in human colon carcinomas and cervical tumour cells [59; 64]. Given the importance of ZNF350/BRCA1-mediated *GADD45* regulation in response to DNA damage, alterations in mRNA expression of these three genes was examined in 116 colorectal cancers. mRNA expression of all three genes was altered with a tendency towards under-expression of *ZNF350* and towards over-expression of *BRCA1* and *GADD45*. Furthermore, a direct correlation between mRNA expression levels of *ZNF350* and *BRCA1*was observed in tumour tissue with respect to normal tissue [64]. The mRNA expression level of *ZNF350* was examined in a panel of cervical cancer cells and clinical cancer specimens and it was found that 7 of 7 cervical cancer cell lines and 9 of 12 cervical cancer specimens expressed significantly reduced amounts of *ZNF350* mRNA when compared with normal cervical tissues [59].

1.3.4 Gene organization and structure

The *ZNF350* gene has been mapped to chromosome 19q13.4, which is located within a cluster of other KRAB-ZFP genes [44]. The *ZNF350* gene spans 2.338 kb and includes 5 exons and 4 introns [65].

Due to the role of ZNF350 in BRCA1-mediated repression of *GADD45*, *ANG1* and *HMGA2*, it may play a role in the occurrence of breast cancer. There have been no germline mutations found in the *ZNF350* gene to date, however some gene variations have been reported. Twelve variants located in the promoter region have been found but do not appear to be statistically associated with breast cancer risk [66; 67]. A further nine polymorphisms have been reported, two within the KRAB domain and one in the second zinc finger yet only one out of 61 patients tested showed a point mutation in the sequence inside one of the polymorphisms and another showing a loss of one allelic variant. This suggests that these type of gene variations and mutations are uncommon in the *ZNF350* gene [63].

1.3.5 Regulation

1.3.5.1 Ubiquitin-Proteasome Pathway

As described above, ZNF350 represses genes involved in cell cycle control, apoptosis and the DNA damage response. However, it is to be assumed that for these genes to be activated in response to DNA damage for example, there must be a way to regulate ZNF350 repression. Following exposure to DNA damage-inducing agents including ultra-violet (UV) and methyl-methanesulfonate (MMS), ZNF350 is ubiquitinated and subsequently rapidly degraded via the ubiquitin-proteasome pathway [68]. Briefly, this process involves the targeting of proteins for degradation by the proteasome through the addition of a polyubiquitin chain (comprising at least 4 ubiquitin molecules) to a lysine residue. This is carried out by three enzymes: ubiquitin-activating (E1), ubiquitin-conjugating (E2) and ubiquitin ligases (E3) with the later being responsible for adding ubiquitin molecules onto the protein and

therefore targeting it for degradation. Since the N-terminal RING finger domain of BRCA1 possesses E3 ligase activity [69], it has been hypothesized that this protein could be responsible for regulating levels of ZNF350. However, BRCA1 does not appear to have any effect on ZNF350 degradation in vivo and an E3 ligase for ZNF350 therefore remains to be identified. The degradation of ZNF350 leads to the derepression of the *GADD45A* gene, allowing for its induction and activity in cell cycle arrest. ZNF350 does not have a degradation signal, an element known to be responsible for the regulated destruction of some proteins. However a 44-amino acid motif, located between the KRAB domain and the eight zinc fingers, has been shown to provide signals for proteasome-mediated degradation. The sequence contains three lysisnes which could potentially bind ubiquitin, and four serines that could be the target of phosphorylation, often the mechanism by which the DNA damage signal is transitted. Cells expressing a ZNF350 mutant lacking this 44-amino acid motif were found to be hypersensitive to DNA damage since ZNF350 could not be degraded and *GADD45A* expression was not fully induced [68].

1.3.5.2 Transcriptional Regulation

The expression of the *ZNF350* gene can itself be repressed by transcription factor binding. The retinoblastoma protein (Rb) is a tumour suppressor that plays an important role in cell cycle progression as well as DNA damage-induced growth arrest. It binds to members of the E2F transcription factor family and represses the expression of E2F-responsive genes involved in the progression of the cell cycle [70]. Interestingly, ZNF350 expression is upregulated in Rb-deficient cells which led to the discovery of a novel E2F1binding motif within the promoter of the *ZNF350* gene. Upon treatment of the DNA damage inducers UV and MMS, the Rb/E2F complex specifically binds to this site and subsequently recruits the CtIP/CtBP complex of chromatin modifiers. This can be correlated with a decrease in ZNF350 promoter activity and consequent derepression of the *GADD45A* gene. The identification of ZNF350 as a downstream target of the Rb/E2F1 complex in

response to DNA damage reveals a potential mechanism of regulation of ZNF350 activity through changes in its level of expression [71].

1.4 IRF-1

Interferon regulatory factor-1 (IRF-1) is the founding member of the interferon regulatory factor family. The human IRF family consists of 10 members that were originally identified as regulators of Type I interferons; however additional roles in a variety of cellular processes have since been discovered, particularly in the antiviral response. Several IRFs play an important role in both innate and adaptive immunity by regulating the development and functions of a variety of different of immune cells [72]. IRFs have also been shown to regulate cellular processes related to oncogenesis which provides a link between the mechanisms involved in both immunity and cancer [73].

All members of the IRF family posses a structurally similar N-terminal DNA binding domain (DBD), which forms a helix-turn-helix structure. The DBDs of all IRFs consist of five conserved tryptophan repeats that are essential for recognising a characteristic IFN-stimulated response element (ISRE) with the consensus sequence 5'-AAN NGA AA-3' [74; 75]. The specific binding site however, may differ between the different members of the family. The C-terminus of the IRF family members, often referred to as the regulatory domain, is less well conserved and may be involved in interactions with other family members, transcription factors or cofactors depending on specific activities of each IRF [76]. A general overview of the structure of the IRF family members can be seen in Figure 1-7.

IRF-1 was first identified as a positive regulator of the *IFNβ* gene by binding to its virus-induced enhancer element. It has subsequently been show to regulate other IFN and IFN-responsive genes and is involved various processes such as the T-cell response, anti-viral response, cell cycle control, DNA damage response, apoptosis and tumour suppression [72; 77-80].

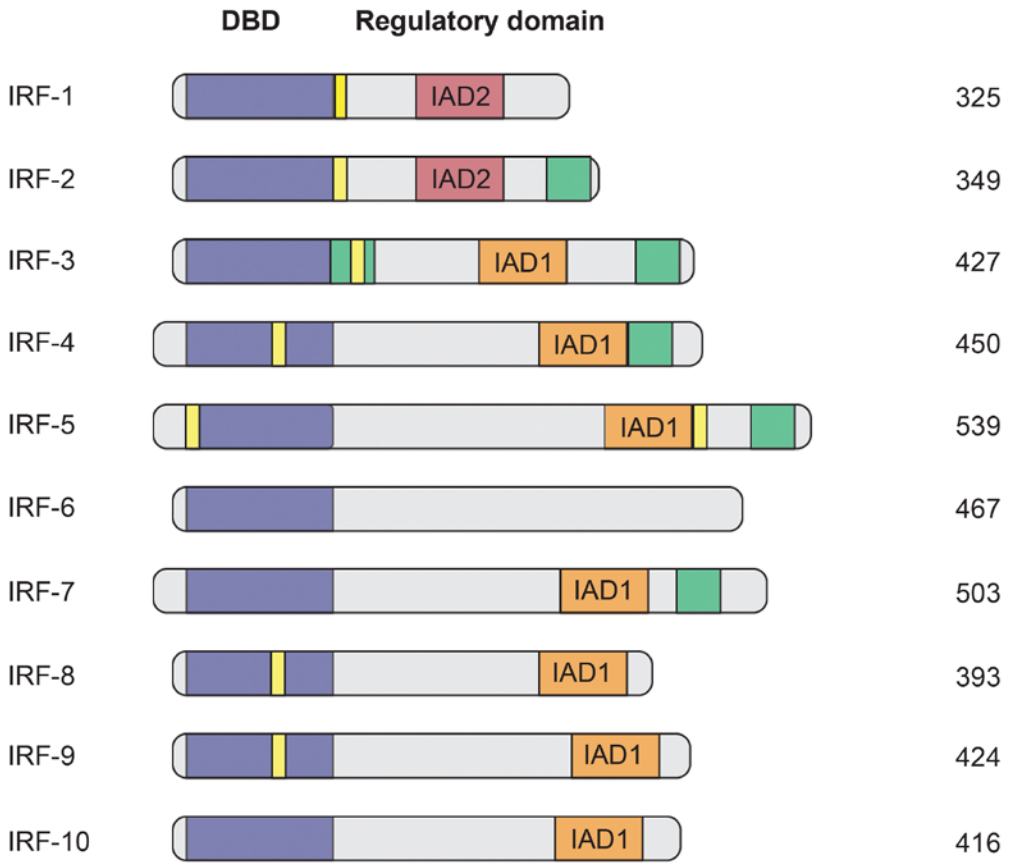


Figure 1-7: Structure of IRF family members

All IRF family members consist of a DNA-binding domain (blue) and a regulatory domain (grey). Most IRFs contain an IRF-association domain of either type 1 (orange) or type 2 (red). Some IRFs also contain a repression domain (green) and/or a nuclear localisation signal (NLS, yellow). The size of each IRF is depicted on the right. Figure adapted from Lohoff and Tak (2005) [81].

1.4.1 Protein Structure

1.4.1.1 DNA binding domain

The DNA-binding domain (DBD) of IRF-1 comprises the first 120 amino acids and is highly conserved throughout the IRF family as well as between species [82]. The DNA binding domain is responsible for interacting with the IRF-E or ISRE on the promoters of IRF-1-response genes, allowing for the activation or repression of gene expression. This interaction is abolished by the occurrence of various point mutations within the IRF-1 DBD, for example when Trp^{11} is mutated to Arg or residues Tyr^{109}, Leu^{112} and Pro^{113} are mutated to Ala [83; 84]. The IRF-1 DBD has been crystallized bound to the positive regulatory domain I (PRD I) found in the *IFNβ* promoter [85]. The structure revealed an α/β architecture with cluster of 3 α helices bordered on one side by a mixed four-stranded β sheet, conferring the characteristic helix-turn-helix motif (Figure 1-8A and B). The α3 helix acts as a recognition helix and selects a core GAAA sequence in the major groove of the DNA and contacts via Arg^{82}, Cys^{83}, Asn^{86} and Ser^{87} which project from the last two turns of the recognition helix. The interaction between IRF-1 DBD and the GAAA motif in the DNA suggests a basis for the occurrence of these repeats in IRF-1 response elements [85].

Despite the crystallized structure of IRF-1 being bound to DNA in monomeric form, footprinting studies by Spink and Evans [86] demonstated that after initially binding as a monomer, two IRF-1 molecules can bind at closely adjacent areas of the inducible nitric-oxide synthase (iNOS) promoter (Figure 1-8C). The synthesis of this gene product is essential for nitric oxide production in a variety of inflammatory conditions. Molecular footprinting has shown that IRF-1 initially binds to a canonical IRF-1 site as a monomer, with a second IRF-1 molecule binding and extending the region in a 5' and 3' direction, with additional contacts in the minor groove of DNA. The binding of the second IRF-1 molecule is dependent on the presence of the initial bound protein, suggesting cooperative binding. Formation of this dimeric complex is necessary for full activation of the iNOS gene by cytokines [86]. A model of the two IRF-1 monomers based on findings from the footpriniting

study, shown in Figure 1-8C, does not suggest that the two monomers have any direct contact. This would imply that the observed binding cooperativity is due to IRF-1-induced conformational changes in the DNA, observed in the crystal structure [85], which increases the affinity of the IRF-1 DBD for the second binding site on the promoter rather than dimerization of IRF-1 molecules. However, since the C-terminal region of IRF-1 is not present in the crystal structure, the possibility that this region may contribute to dimer formation cannot be ruled out.

1.4.1.2 Mf2 Domain

The multi-functional domain 2 (Mf2) of IRF-1 has been mapped to residues 106-140 and is thought to act a binding interface for a variety of different proteins [87]. These interacting proteins include the IRF-1 E3 ligase CHIP, YB-1, NPM1 and KAP1, which all have different amino acid specificities [87; 88]. Recently, the E3 ligase MDM2 has been found to bind to the Mf2 domain of IRF-1 and ubiquitinate residues in the DBD. If MDM2 cannot interact with IRF-1, for example when it is bound to DNA, ubiquitination cannot occur suggesting a way in which active IRF-1 could be protected from degradation [89]. The sequence of the Mf2 domain suggests that it is highly disordered which confers a flexible structure and allows the binding of a variety of diverse proteins. This type of intrinsically disordered domain appears to be common in proteins and transcription factors involved in cell signalling pathways as it allows for multiple, transient interactions [87].

1.4.1.3 Nuclear localisation signal

The nuclear localisation signal (NLS) lies mostly within the Mf2 domain (aa 117-141) and is usually found to be exposed and easily accessible. Since IRF-1 is a transcription factor it is mainly found in the nucleus of cells enabling it to bind DNA and regulate transcription. When the NLS is deleted, IRF-1 is localised mostly in the cytosol, rendering it inactive as a transcriptional regulator [82]

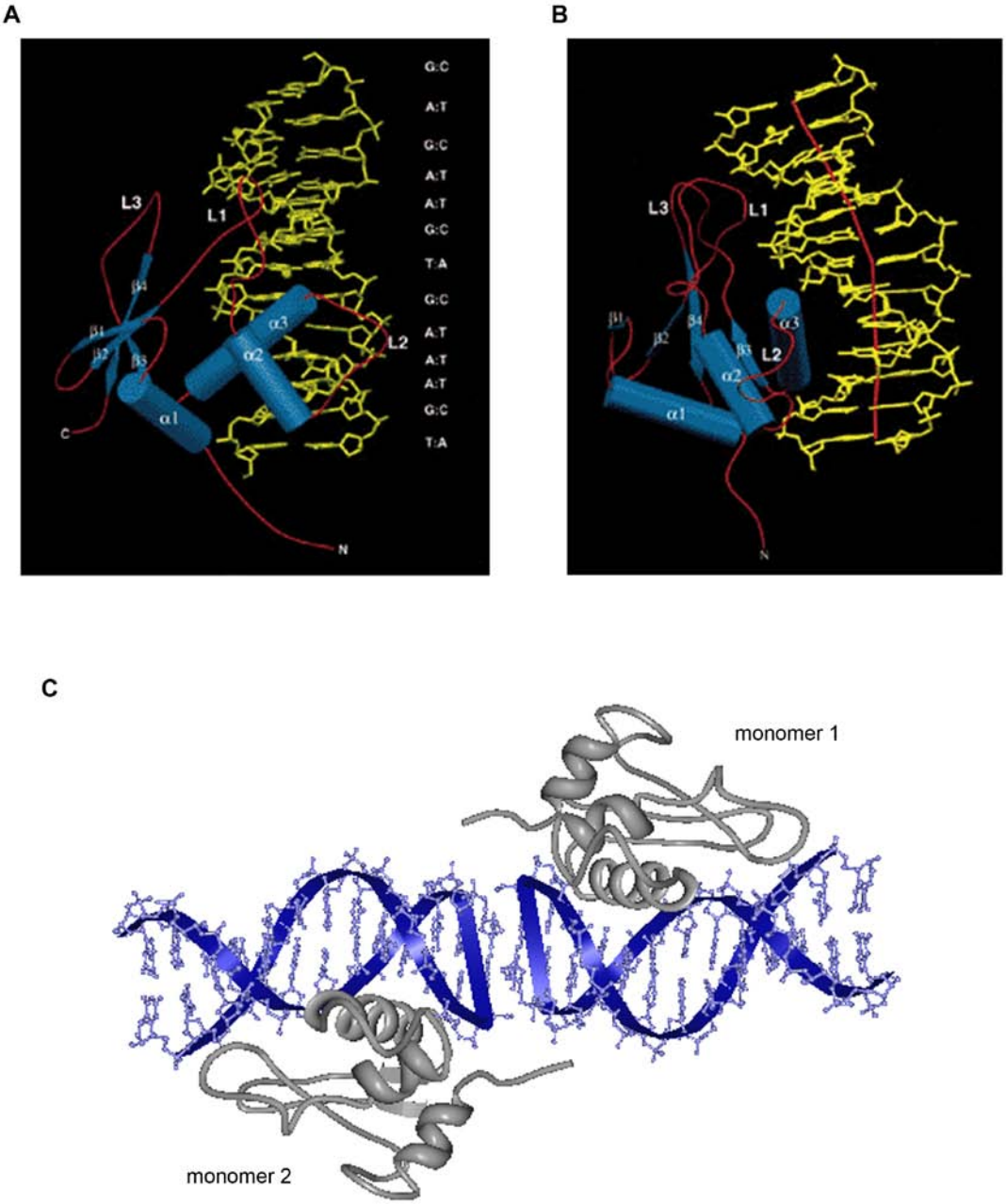


Figure 1-8: Overview of IRF-1 bound to DNA

(A) The α/β architecture of the IRF-1 DNA-binding region is interrupted by multiple loops labelled L1, L2 and L3 that flare out and contact the DNA backbone. (B) A view along the recognition helix α3 showing the bending of the DNA towards the protein. The local helical axis was calculated with the program CURVES and is shown as a solid curve within the DNA. Both images in (A) and (B) from Escalante *et al.* [85]. (C) Model of cooperative binding of two IRF-1 monomers to the iNOS promoter (model based on the crystal structure and data from a study by Spink *et al.*) [86].

1.4.1.4 Transactivation domain

The transactivation domain of IRF-1 was first mapped to the C-terminus by Lin et al [90]. The study speculated that two regions of acidic amino acid stretches (226-230 and 276-281) could play a role in transactivation since similar regions were found to in the transactivation domains of NF-κB p65 subunit and yeast GAL4 proteins [90-92]. Further studies revealed in fact that a region between residues 185-256 was sufficient for transactivation activity and when this fragment was fused to GAL4 in a CAT-reporter assay, expression was significantly increased compared to GAL4 alone [82; 93]. Closer analysis revealed residues 217-260 are required for sufficient transactivation with the core activation domain comprising residues 233-255. Interestingly, mutants lacking this core domain function as a dominant negative in cells analogous to the activity of IRF-2, and efficiently inhibit IRF-1 dependent cell cycle regulation and apoptosis [94].

1.4.1.5 Dimerisation domains

IRF-8 has been shown to repress IRF-1-mediated transcriptional activation of IFN-inducible genes by heterodimerisation [95; 96]. By using deletion mutants, the binding interface between IRF-8 and IRF-1 was established to be in a region encompassing residues 164-219. In addition, when part of the NLS (aa 124-141) was mutated, IRF-8 interaction with IRF-1 was perturbed, suggesting this region must contribute to heterodimerization [82].

A study with murine IRF-1 has shown its ability to homodimerise *in vivo* through residues 91-114 [97]. However, *in vitro* IRF-1 is present in solution as a monomer, as shown by native PAGE and density gradient centrifugation (unpublished data, Ball group). In addition, IRF-1 was crystallised in its monomeric form [85] but as mentioned earlier, footprinting studies have suggested that IRF-1 molecules cooperatively bind at the iNOS promoter [86], possibly due to an additional homodimerisation domain outwith the IRF-1 DBD.

1.4.1.6 Enhancer domain

The C-terminal region of IRF-1 (aa 256-325) has the ability to 'enhance' its own transactivation despite the fact that it does not possess any intrinsic transactivation activity. This region is known to function in various capacities including protein binding, degradation, growth regulation and repression.

There are two sites (aa 226-245 and aa 271-290) in the enhancer domain that have been shown to directly bind the co-activator p300. This leads to the acetylation of p53 at the p21 promoter and subsequent activation of gene expression. Further analysis revealed that the DNA-binding activity of IRF-1 was not necessary and only its direct interaction with p300 was required for p21 gene regulation [84].

When the enhancer domain is mutated, the ability of IRF-1 to suppress cell growth is impeded [98]. Studies into the C-terminal 25 aa of IRF-1, also known as the multi-functional domain 1 (Mf1), have revealed that this region may play a role in the repression of the *Cdk2* gene, the product of which is a component of cell cycle control machinery [99]. Within this sub-domain, 12 residues (301-313), including an LXXLL co-signature motif, appear to be critical, since mutations within the motif result in a loss of *Cdk2* repression and ultimately growth suppression [98]. Interestingly, LXXLL motifs are known to participate in many protein-protein interactions associated with different aspects of transcriptional regulation [100]. The C-terminal 12 residues also appear to contain a negative regulatory signal for IRF-1-dependent *IFNβ* since the deletion of this region results in an increase in IRF-1 activation of the *IFNβ* promoter [98].

Studies into the degradation of IRF-1 revealed the presence of discrete motifs in the enhancer domain were involved in polyubiquitination and degradation [101]. This region of IRF-1 is not ubiquitinated, however removal of the last 70 amino acids of the C-terminus inhibits both polyubiquitination and degradation. Interestingly, when only the last 25 amino acids are removed only the degradation, and not ubiquitination, of IRF-1 is inhibited and further analysis of this region mapped the

IRF-1 degron to residues 301-310, a region which partially overlaps with the *Cdk2* repressor motif [101].

Additionally, the last 25 amino acids of the IRF-1 C-terminus have been identified as a target for a monoclonal nanobody. Nanobody binding to this region results in up to an 8-fold increase in activity of IRF-1-resonsive promoters as well as an increase in the rate of IRF-1 degradation [102]. This surprising correlation has been seen with other transcription factors where they appear to be most potent when their levels are reduced. This mechanism may suggest a possible direct role for the ubiquitin-proteasome pathway in transcriptional activation [103].

These studies suggest that the IRF-1 enhancer domain may be an interface for protein interactions, possibly binding co-repressor proteins to mediate *Cdk2* repression and components of the ubiquitin-pathway such as an E3 ligase, as well as binding p300.

1.4.2 Gene Expression

The IRF-1 gene has been mapped to the long arm of chromosome 5 (5q31.1), which is a region frequently deleted in many types of leukemia [104]. Analysis of the *IRF-1* promoter has revealed that it does not possess a TATA box but instead contains a putative CAAT box which is flanked on either side by NF-κB or κB motifs [105; 106]. NF-κB plays a key role in regulating the immune response to infection and has been linked to virus or double-stranded RNA-induced *IFNβ* gene regulation [107; 108]. Interestingly, when a point mutation is introduced within the NF-κB binding site on the *IRF-1* promoter, binding is abolished and virus inducibility is considerably reduced in L929 cells. This would suggest that NF-κB or NF-κB-like factors may be involved in *IRF-1* induction upon viral infection [106].

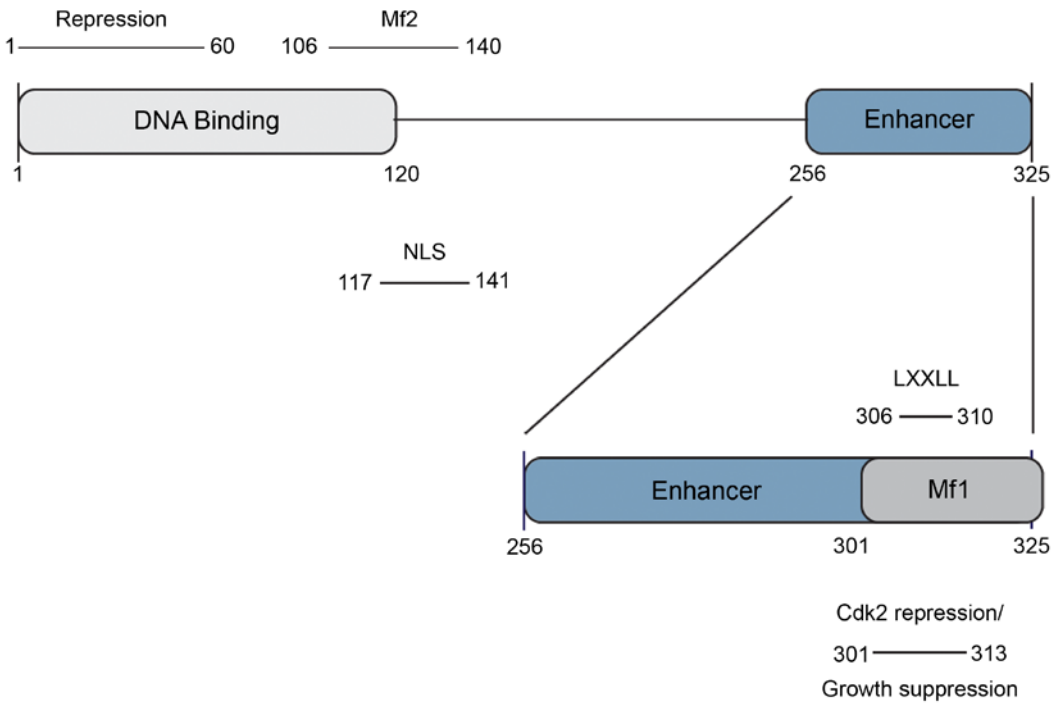


Figure 1-9: Functional domains of the human IRF-1 protein

NLS nuclear localisation signal; *Mf1* multifunctional domain1; *Mf2* multifunctional domain2; *LXXLL* is a co-signature motif.

1.4.2.1 JAK-STAT pathway

IRF-1 gene expression can be induced by virus, double-stranded RNA, interferons (IFNs) and other cytokines and genotoxic stress [109; 110]. IRF-1 can then go on to regulate the expression of various target genes involved in a wide variety of cellular processes [111]. *IRF-1* induction is mainly through the JAK-STAT pathway which is triggered by IFN-γ, a cytokine produced by activated T cells or NK cells. IFN-γ binds to the transmembrane IFN-γ receptor or IFNGR which leads to the activation of receptor-associated Janus protein kinase 1 or 2 (JAK1 or JAK2) through auto- and/or trans- phosphorylation of specific residues. Activated JAKs then phosphorylate the downstream targets, signal transducers and activators of transcription (STAT) 1 and 2. This results in the activation of the STATs which in turn triggers the formation of the transcriptional activator complex IFN-γ-activated factor (GAF), also known as IFN-α-activated factor (AAF) [112; 113]. GAF is a homodimer of activated STAT1 which then moves into the nucleus and binds to specific DNA elements known as IFN-γ-activated sequence, or GAS. This results in the transcriptional activation of various IFN-stimulated genes, including IRF-1 [113].

1.4.2.2 ATM pathway

The role of IRF-1 in response to viral attack is well established, however less is known about its induction and role in response to other environmental stresses, such as DNA damage. Eukaryotic cells respond to DNA damage by activating damage checkpoint pathways, which lead to the arrest of cell cycle progression and alter the expression of genes to allow for repair and/or apoptosis. It has been proposed that IRF-1 is activated in response to ionising radiation (IR) [111; 114] and a study by Pamment et al. [109] has suggested a role for IRF-1 in regulating cell growth not only as a result of IR exposure but also in response to DNA strand breaks caused by the inhibition of topoisomerase activity. The study found that IRF-1 induction following genotoxic stress is regulated by ATM-signalling events, particularly ATM kinase, resulting in changes in IRF-1 mRNA levels and protein synthesis.

Furthermore, an increase in IRF-1 protein levels and post-translation changes in the rate of protein degradation were also observed. Interestingly, in ATM-deficient cells, IRF-1 protein and mRNA induction in response to genotoxic stress is inhibited however, induction in response to viral mimetics remains unaffected [109]. This would suggest that IRF-1 is induced by two distinct pathways; the ATM-signalling pathway in response to DNA damage and the JAK/STAT-signalling pathway in response to viral infection [109; 112; 113].

1.4.3 Roles of IRF-1

1.4.3.1 Immune System

IRF-1 has been found to function extensively in the immune response with roles involved in inflammation, NK cell development, T cell differentiation and the innate immune response [73]. The innate immune response depends on the recognition of pathogen-associated molecular patters (PAMPs), such as lipopolysaccaride (LPS) and viral nucleic acids, by pattern recognition receptors (PRRs). These receptors fall into two categories, cytosolic PRRs and toll-like receptors (TLRs). The binding of PAMPs to PRRs activates signal transduction pathways and induces various target genes such as type I IFNs, proinflammatory cytokines and chemokines, thereby eliciting antimicrobial responses [115]. IRF-1 has been shown to induce two cytosolic PRRs, *RIG-1* (retinoic acid inducible gene 1) and *PKR* (protein kinase, dsRNA dependent) as well as enhancing TLR-mediated gene induction [116-118].

1.4.3.2 Autoimmunity

Studies in IRF-1-deficient mice have revealed that IRF-1 regulates the induction of molecules with important roles in inflammation, such as inducible nitric oxide synthase (*iNOS*) and interleukin-1β-converting enzyme (*ICE*) [119-121]. Further analysis has shown that IRF-$1^{-/-}$ mice exhibit significantly decreased incidence and severity of inflammatory disorders such as type II collagen-induced arthritis (CIA)

and experimental allergic encephalomyelitis (EAE) [122]. IRF-1 has been found to repress the Secretory leukocyte protease inhibitor (*SLP1*) gene, which plays a role in neutrophil-mediated inflammation and the LPS response [123], whilst it activates Cyclooxegenase-2 (*Cox-2*), which is involved in the formation of prostanoids, biological mediators of inflammation [124]. In addition, IRF-1 in cooperation with NF-κB activates Vascular cell adhesion molecule-1 (*VCAM-1*), a protein involved in recruiting inflammatory cells from the bloodstream to the site of tissue injury [125].

1.4.3.2.1 HIV-1

Many viruses are able to evade the immune system by inhibiting cellular IRFs. The most fascinating mechanism of immune evasion is that used by human immunodeficiency virus (HIV-1). The infection of cell by viruses, including HIV-1, induces the expression of IRF-1. However, the long terminal repeat (LTR) of *HIV-1* contains an ISRE to which IRF-1 and IRF-2 can bind and activate gene expression. Therefore, when HIV-1 infects the cell, the increased production of IRF-1 activates *HIV-1* gene transcription and enables its replication. In addition, IRF-2 which usually inhibits IRF-1 activity, appears to act as an agonist and ultimately assists in the replication of HIV-1 [126].

Altered susceptibility to HIV-1 infection has been observed in multiple cohort studies, with one of the best characterized HIV-1-exposed, uninfected group being a commercial sex worker cohort from Nairobi, Kenya [127; 128]. Analysis of the IRF-1 gene identified three polymorphisms which showed associations with resistance to HIV-1 infection. Peripheral blood mononuclear cells (PBMCs) from patients with protective IRF-1 genotypes displayed significantly lower basal IRF-1 expression and reduced responsiveness to IFNγ stimulation [129]. Further studies in these cells also demonstrated significantly lower HIV-1 LTR transcription during the initial stages of infection compared to PBMCs with other haplotypes, which correlated with the kinetics of IRF-1 responsiveness to HIV-1 infection in cells. This suggests that IRF-1 genotypes can alter early HIV-1 LTR transcription, possibly by modulating the expression of IRF-1 [130].

1.4.3.3 Human Cancer

A number of clinical studies have indicated that a loss of *IRF-1* is linked to the development of some forms of cancer in humans, in particular leukaemia. The chromosomal region 5q31.1, the region where the *IRF-1* gene is located, frequently shows abnormalities in leukaemia and preleukaemic myelodysplastic syndrome (MDS) [104]. It has been reported that among 13 patients with leukaemia or MDS, *IRF-1* was the only gene found to be consistently deleted or rearranged on one or both alleles and in patients where MDS has progressed to acute myeloid leukaemia (AML), 12 out of 14 were found to have a loss of one allele of *IRF-1* [104; 131]. This loss of a single *IRF-1* allele has also been found in oesophageal and gastric cancers, with one in four cases of gastric cancers showing this deletion is accompanied by an inactivating point mutation in the other allele [132-134]. In addition, decreased expression of IRF-1 mRNA has been reported in several types of cancer such as CML, endometrial cancer and hepatocellular carcinoma with low levels of IRF-1 correlating with metastasis and invasive potential in breast cancer cell lines [73; 135].

In addition to genetic alterations of the *IRF-1* gene, several mechanisms can lead to the loss-of-function of IRF-1. Nucleophosmin (NPM) is a nuclear phosphoprotein, chaperone and putative ribosome assembly factor that has been implicated in the regulation of a number of transcription factors. It can act as an oncogene and is often overexpressed in leukemic cells. NPM has been shown to bind directly to IRF-1 and prevent its DNA binding and transcriptional activity [136]. IRF-1-induced transcriptional activation of *IFN* genes can be inactivated by the oncoprotein E7 of human papilloma virus (HPV) which recruits HDACs to promoters, whilst IRF-2, a structurally similar protein to IRF-1, can act as an antagonist and bind to the same regulatory elements [78; 137].

Kaposi's sarcoma-associated herpes virus (KSHV; also known as human herpes virus 8, or HHV8) is a human DNA tumour virus that causes Kaposi's sarcoma, primary effusion lymphoma and some cases of multicentric Castleman's disease, which is a rare disease of the lymph nodes causing over proliferation of lymph cells.

(MCD) [138-140]. To assist in its infection of cells, the KSHV genome encodes many immunomodulatory proteins that act by targeting different aspects of the human adaptive and innate immune responses. These proteins include 4 viral interferon regulatory factors (vIRF-1 to 4), which are homologous to cellular IRFs and act by inhibiting the IRFS and consequently IFN signalling [141; 142]. vIRF-1 acts by associating with the histone acetyltransferase p300 and prevents the formation of the CBP/p300:IRF-3 complex, thus resulting in defective IRF-3-induced IFNα production [143; 144].

1.4.3.4 Tumour suppressor

The tumour suppressor-like activity of IRF-1 has been demonstrated using *IRF-1*$^{-/-}$ MEFs. The malignant transformation of normal MEFs requires the activities of at least two oncogenes however, the loss of IRF1 and the introduction of c-Ha-ras is sufficient to transform *IRF-1*$^{-/-}$ MEFs [145]. In addition, cell growth inhibition in myeloid cells by activated Ras involves IRF-1 and the induction of $p21^{WAF1/CIP1}$ [146]. Multiple studies have also shown that ectopic expression of IRF-1 can suppress the malignant properties of cancer cell lines in vitro and in vivo [73]. In mouse studies, the loss of IRF-1 alone does not induce tumour formation, however in combination with over-expression of c-Ha-ras or loss of p53 the occurrence of tumours increases substantially [79]

1.4.3.5 Invasion and metastasis

Enhanced expression of matrix metalloproteinase-9 (MMP-9) correlates with invasion during tumour progression [147]. Tumour necrosis factor-α (TNF-α) can activate the *MMP-9* gene through NF-κB while IFNs can inhibit this activation (references in [148]). The *MMP-9* promoter contains an NF-κB consensus sequence with an overlapping IFN-responsive-like element (IRE). It has been shown that IRF-1 transfection can block TNF-α-mediated MMP-9 activation by competing with NF-

κB for binding to the *MMP-9* promoter. This observation is particularly interesting since IRF-1 and NF-κB usually act synergistically but in this case IRF-1 acts as competitive inhibitor [148].

1.4.3.6 Cell cycle

The expression of IRF-1 is regulated throughout the cell cycle as shown in a study where serum starved NIH3T3 cells had significantly elevated levels of IRF-1 mRNA, but when cell cycle progression is restarted by the addition of serum, IRF-1 mRNA levels rapidly reduce [149]. IRF-$1^{-/-}$ MEFs are deficient in their ability to undergo DNA damage-induced cell cycle arrest, a phenotype similar to that seen in $p53^{-/-}$ MEFs [114]. When wild-type MEFs are exposed to DNA damaging agents such as γ-irradiation they undergo G1 cell cycle arrest due to the upregulation of the $p21^{WAF1/CIP1}$ CDK inhibitor. This has been shown to be dependent on both IRF-1 and p53 and while the $p21^{WAF1/CIP1}$ promoter contains both IRF-1 and p53 response elements, IRF-1 is only a weak activator of $p21^{WAF1/CIP1}$ expression whereas p53 can induce expression several fold [114]. Studies have shown that in response to DNA damage, there is an increase in IRF-1mRNA expression and protein half-life in an ATM-dependent manner. This results in elevated IRF-1 protein levels which cooperates synergistically with p53 to induce $p21^{WAF1/CIP1}$ expression [109; 114]. Further studies by the Ball group demonstrated that IRF-1 recruits the co-activator p300 to p53-bound DNA thereby up-regulating $p21^{WAF1/CIP1}$ expression [84].

1.4.3.7 Apoptosis

Apoptosis is a key mechanism by which cancerous cells can be eliminated from the host and IRF-1 has been shown to induce apoptosis in response to DNA damage or other stimuli. Apoptosis occurs through two pathways, an intrinsic one initiated by the release of factors from the mitochondria, or an extrinsic pathway initiated by the stimulation of transmembrane death receptors. When wild-type MEFs are introduced

to an activated oncogene such as c-Ha-ras, instead of arresting the cell cycle they undergo apoptosis when treated with anti-cancer drugs or DNA damage-inducing agents. This apoptosis is characteristic of tumour suppression and is dependent on both IRF-1 and p53 [145].

In mitogenically activated T lymphocytes, DNA damage-induced apoptosis is dependent on IRF-1 but independent of p53 whereas the opposite is seen in thymocytes where it is dependent on p53 but not IRF-1 [120; 150]. This reveals that IRF-1 and p53 can regulate DNA damage-induced apoptosis both cooperatively and independently depending on the type of cell [114]. A noteworthy point is that p53 and IRF-1 can both be transcriptionally activated by the Gate keeper of apoptosis activating protein (GAAP-1) which has pro-apoptotic activity [151].

IRF-1 has been linked to apoptosis in a variety of cell types and so far several IRF-1 target genes involved have been identified. In breast cancer MDA468 cells, caspase-8 was found to play a key role in IRF-1-dependent apoptosis while in HIV-infected macrophages, regulation of TNF-related apoptosis-inducing ligand (TRAIL) involved IRF-1 along with STAT1 and IRF-7 [152; 153]. In addition, in gastric cancer cells, IRF-1 was shown to transcriptionally activate the p53-upregulated modulator apoptosis (PUMA) in the absence of p53 and initiate extrinsic pathway of apoptosis through PUMA [154]. IRF-1 can also activate X-linked inhibitor of apoptosis protein-associated factor-1 (*XAF-1*) which is a gene encoding a tumour suppressor that sensitises cancer cells to apoptosis [155]. Evidently, IRF-1 can act to induce apoptosis in cells by a variety of different mechanisms.

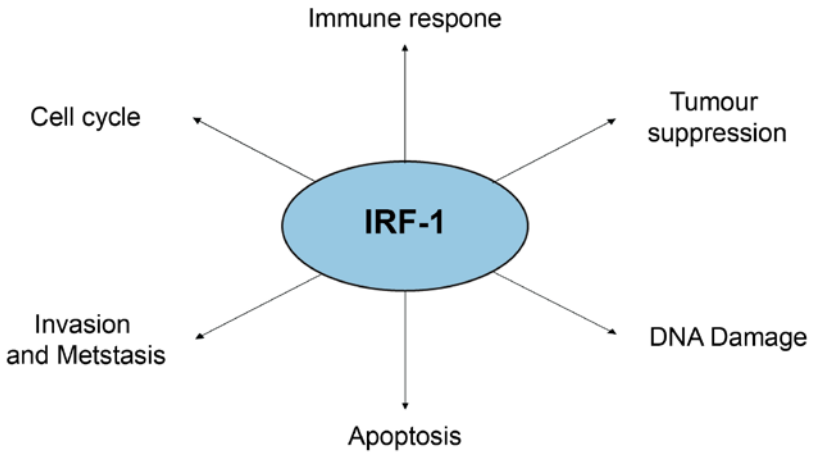


Figure 1-10: Schematic summarising the multiple functions of IRF-1

1.5 Objective of this thesis

ZNF350 has been implicated in several important cellular pathways including invasion and metastasis and the DNA damage response. In order to better understand how ZNF350 functions, I sought to explore three main areas. Firstly, do IRF-1 and ZNF350 interact in vivo and is this interaction affected by different cellular signals? This is of interest to help in understanding the role these two proteins play in the cell, for example in response to immune system activation and DNA damage. Since both IRF-1 and ZNF350 have been implicated in transcriptional regulation, the next question I sought to answer was what genes do they bind to and do they have any common targets? Often in cancer gene expression is dysregulated, so by identifying IRF-1 and ZNF350 targets a greater understanding can be gained on how they act to control gene expression. This could aid in the development of therapies to counteract aberrant gene expression. Gene transcription is often regulated by multi-protein complexes so the final question I aimed to answer was what other proteins does ZNF350 bind to? By answering these three questions, it is hoped that the relationship between ZNF350 and IRF-1 and the role they play in controlling gene expression is better understood.

Chapter 2: Materials and Methods

2.1 Reagents, Plasmids and centrifuges

pcDNA3-IRF-1 wt was kindly gifted from Vikram Narayan, Emmanuelle Pion and Kathryn Ball; One-strep IRF-1 wt from Vikram Narayan; pET-15b-ZNF350 from Angeli Möller and pLTR-Luc-HIV-1 was very kindly gifted from Nishitsuji et al. (2012) [56].

All general chemicals and reagents were from Sigma or BDH and for centrifugations, Eppendorf 5415R and Sorvall RC-5C plus were used unless otherwise stated.

2.2 General microbiological techniques

2.2.1 Maintaining bacterial cultures

Bacterial cultures were grown in Luria-Bertani (LB) broth with 100 µg/ml ampicillin (SIGMA) with inoculation from either a glycerol stock or a bacterial culture dish. The culture was incubated in an incubator-shaker at 37^{o}C and 220 rpm.

Luria-Bertani (LB) broth

25 g of LB medium (Miller) was dissolved in 1 litre of distilled water and sterilised by autoclaving at 121^{o}C for 15-20 min. Final concentrations of the individual components in the broth were:

1% (w/v) Tryptone
1% (w/v) NaCl
0.5% (w/v) Yeast extract

2.2.2 Glycerol stocks

For long term storage of the bacteria cultures, glycerol stocks were prepared by adding 200 µl sterile glycerol to 800 µl mid-log phase bacterial culture in a cryotube (Nunc). The stocks were mixed by gentle agitation, snap frozen and stored at -80°C.

2.2.3 Preparation of competent cells

Bacterial cells from glycerol stock were inoculated into 2 ml of LB media (without antibiotic) and incubated overnight at 37°C in a shaking incubator at 220 rpm. An aliquot of the overnight culture (250 µl) was added to 50 ml LB media and grown until the OD_{600nm} reached approximately 0.4. The culture was centrifuged at 4000 g for 15 min at 4^{o}C and the pellet was re-suspended in 16 ml ice-cold Buffer I. Following incubation for 10 min on ice, cultures were centrifuged again as above and the pellet was re-suspended in 2 ml ice-cold buffer II. The cell pellet was then incubated for a further 10 min on ice and aliquoted (30 µl) and snap frozen using liquid nitrogen. Competent cells were stored at -80°C.

Buffer I

60 mM CH_3COOK
10 mM $CaCl_2$
40 mM $MgCl_2$
100 mM RbCl
15% (v/v) glycerol
Adjusted to pH 5.8 with CH_3COOH and sterilised by filtration

Buffer II

10 mM MOPS
75 mM $CaCl_2$
10 mM RbCl
15% (v/v) glycerol

Adjusted to pH 6.5 with NaOH and sterilised by filtration

2.2.4 Transforming bacterial cells

Plasmid DNA (usually 250 ng, or between 50-250 ng) was mixed with an aliquot of competent cells (30 μl) and incubated for 30 min on ice to thaw. The bacteria were the heat- shocked at 42°C for 45 seconds in a water bath and cooled on ice for 2 min. LB broth without antibiotic (500 μl) was then added and the mixture was incubated for 90 min at 37°C with shaking. Aliquots (100 μl and 250 μl) of cells were plated on agar plates and incubated at 37°C overnight.

LB Agar

40 g LB-Agar (Miller) was dissolved in 1 litre of distilled water and sterilised by autoclaving at 121^{o}C for 15-20 min. Final concentrations of the individual components were:

1% (w/v) Tryptone
1% (w/v) NaCl
0.5% (w/v) Yeast extract
1.5% (w/v) Agar

LB-agar plates were prepared by pouring liquefied LB-agar, produced by heating ans subsequently cooling to about 40^{o}C, into 90 mm petridishes (Sterilin). If required, ampicillin was added to the liquefied LB-agar immediately prior to pouring into the petridishes. The agar was further cooled until it solidified, and the plates were dried at 37^{o}C for upto 1 h prior to use.

2.2.5 Plasmid DNA amplification, extraction and quantification

A single colony from a LB agar plate was inoculated in 5 ml (Mini-prep) or 250 ml (Maxi-prep) of LB broth containing ampicillin and incubated for 12-16 hours at 37°C with shaking (220 rpm). Cells were collected by centrifugation at 4000 rpm for 20 min at 4^{o}C and plasmid DNA was extracted using either the Qiagen HiSpeed Mini-prep or Maxi-prep kit following supplier's instructions. DNA was eluted using 30 or 500 µl nuclease-free water, respectively and stored at -20^{o}C. DNA content was quantified using a NanoDrop ND-1000 spectrophotometer following manufacturer's instructions.

2.2.6 Agarose gel electrophoresis

DNA samples were analysed using agarose gel electrophoresis. 1- 2% agarose gels were prepared as required by dissolving electrophoresis-grade agarose (Invitrogen) in 1X TAE by gentle heating in a microwave. After all the agarose had dissolved, the mixture was allowed to cool down slightly and SYBR® Safe DNA gel stain (life technologies) at 1:10,000 dilution was added to the gel prior to pouring into a gel chamber. DNA samples were mixed with 6X DNA loading dye at a 5:1 ration of sample:dye and loaded onto the gel which was subsequently run at 75-100 V for 45-60 minutes.

1X TAE Buffer

40 mM Tris-HCl
1 mM EDTA
Adjusted pH to 8 with glacial acetic acid

6X DNA Loading Dye

0.25% bromophenol blue
0.25% xylene cyanol
15% glycerol

2.2.7 DNA sequencing

DNA sequencing was carried out using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) according to the manufacturer's instructions. The internal sequencing primers used for sequencing for IRF-1, ZNF350 and MMP9 as well as any commercially available sequencing primers used in this thesis are listed in Table 2-1.

Table 2-1: Sequencing Primers

Primer	***Sequence***	***Forward/ Reverse***
ZNF350F1	GGAATCCATAACACTGGAGG	Forward
ZNF350R1	GGGACTCCTTTCTTCATGC	Reverse
ZNF350F2	CATCAGTCCCAAGCATCAG	Forward
ZNF350R2	CGTGTCTCATAGCACATCAGAG	Reverse
ZNF350F3	CACACAGGAGAGAAACCC	Forward
ZNF350R3	GCAGTGAATGTGGTTGTG	Reverse
IRF-1 S1	GGAGCCAGATCCCAAGACGTG	Forward
IRF-1 S6	GGTGGCATCCATGTTCTTCAG	Reverse
MMP9F1L	GGGGTACCAGTGACTTGCCCAAGGTCACATAGC	Forward
MMP9F1S	GGGGTACCAGTGACTTGC	Forward
MMP9R1	GCTGAGGCAGGAGAATTG	Reverse
MMP9F2	GTGACAGAGTGATACTACACCC	Forward
MMP9R2	GGCTCACTGTATCCTTGACC	Reverse
MMP9F3	GAATGAGCCACCATACCTG	Forward
MMP9R3	GGAAGCTGAGTCAAAGAAGG	Reverse

MMP9F4	GGAGTCTTCCATCACTTTCC	Forward
MMP9R4L	CCTGGAGACCTGAGAACCAATCTCAAAGCTTGGG	Reverse
MMP9R4S	CCAATCTCAAAGCTTGGG	Reverse

Sequencing reactions were set up as follows:

Sequencing PCR

1 µl Big Dye v3.1
2 µl Big Dye buffer (5X)
300 ng DNA template
0.5 µl sequencing primer (3.2 µM stock)
Nuclease-free water to 10 µl

PCR reactions were set up as described above using each sequencing primer individually.

Thermal cycling conditions:

96^{o}C for 1 min
96^{o}C for 10 sec
50^{o}C for 5 sec
60^{o}C for 4 min
Repeat steps 2-4 for 25 cycles
Hold at 4^{o}C

2.2.8 Ethanol/EDTA precipitation

After thermal cycling, EDTA (2.5 µL of 125 mM stock) and 100% ethanol (30 µl) were added to each reaction mix. The mix was gently vortexed for 5 sec and incubated at room temperature for 20 min. Samples were centrifuged at 4000 rpm for 20 min at room temperature and the supernatant was removed gently to avoid disruption of the cell pellet. The samples were centrifuged for a further 30 sec and any remaining liquid was discarded. The DNA pellet was then rinsed with 30 µl of 70% ethanol and centrifuged at 4000 rpm for 5 min and the remaining ethanol was subsequently removed. The cell pellet was air dried for 15 min and submitted for sequence analysis to the Source Bioscience DNA Sequencing Service at Cambridge

2.2.9 Amplification of gene by PCR

Template DNA was used to amplify the desired gene of interest. Primers were designed to incorporate different restriction enzyme (RE) sites into the 5' and 3' end of the insert. A list of primers with RE sites used in this thesis is given in Table 2-2.

Table 2-2: Primers used for cloning

Gene to be cloned	*Target Vector*	*Primers (5' to 3')*	*DNA template*
ZNF350	pCOLDI	Forward: GACGTACATATGATCCAGGCCCAGGAATC Reverse: GCCGGCGGATCCCTATGGGTTTTCTGTAACATAAA ATAAG	pET-15b-ZNF350
MMP9	pGL3	Forward: GGGGTACCAGTGACTTGCCCAAGGTCACATAGC Reverse: CCCAAGCTTTGAGATTGGTTCTCAGGTCTCCAGG	Human genomic DNA

Key:

Blue = Restriction enzyme recognition sequence

Red = Linker sequence

Black = ZNF350 or MMP9 sequence

Table 2-3: Sequences of Restriction Enzymes

Restriction Enzyme	***Recognition sequence (5' to 3')***
CATATG	NdeI
GGATCC	BamHI
GGTACC	KpnI
AAGCTT	HindIII

The PCR reaction mix using 2X Pfu Master Mix (Rovalab) was prepared as follows:

12.5 µl 2X Pfu Master Mix

2.5 µl Band Doctor

50 ng Template DNA

0.13 µl Forward Primer (100 µM stock)

0.13 µl Reverse Primer (100 µM stock)

Nuclease-free water to 25 µl

Thermal cycling conditions:

95°C for 2 min

95°C for 20 sec

58°C for 40 sec

72°C for 2 min

Repeat steps 2-4 for 25-30 cycles

72°C for 5 min

Hold at 4°C

Alternatively, for the amplification of -1940 MMP9 DNA fragment, the PCR reaction was carried out using Taq Polymerase. The reaction mix was prepared as follows:

Taq polymerase

1.25 µl Taq Polymerase
1 µl dNTPs (10 mM stock)
2.5 µl Forward Primer (0.5 µM stock)
2.5 µl Reverse Primer (0.5 µM stock)
30 ng Genomic DNA
10X Taq Buffer
5X Band Doctor
Nuclease free water to 50 µl

Thermal cycling conditions:

95°C for 2 min
95°C for 30 sec
70°C for 30 sec
72°C for 2 min
Repeat steps 2-4 for 35 cycles
72°C for 5 minutes
Hold at 4°C

Following PCR, the amplified DNA was cleaned up using the Qiagen PCR Clean-up Kit and eluted in 30 µl of nuclease free water. A 2 µl aliquot of this was loaded on a 1% agarose gel to check for successful amplification.

(I) Restriction Digest of vector and insert

Double digests were set up for the vector and insert using REs purchased from New England Biolabs (NEB), as indicated below. NEB buffers were also used and double digest conditions were as recommended by the supplier.

~30 µl PCR product (or 3-4 µg vector)
4 µl NEB Buffer
0.4 µl BSA (if required)
1 µl RE 1
1 µl RE 2
Nuclease free water to 40 µl

Double digests were carried out at 37°C for 90 min followed by incubation at 65°C for 10 min to deactivate the REs. Following the double digest, the entire PCR mix was loaded onto a 1% agarose gel and single bands corresponding to the double digested vector or insert were cut out and purified using the Qiagen Gel Extraction Kit, according to the manufacturer's instructions.

(II) Ligation of double-digested vector and insert

Ligation of the gel-extracted insert and vector double-digests was carried out using T4 DNA ligase (Promega) according to the manufacturer's instructions. 100 ng of vector was used and the amount of insert to be added was calculated using the following formula:

$$Insert_{ng} = \frac{Vector_{ng} \times Insert\ size_{kb}}{Vector\ Size_{kb}} \times\ Molar\ ratio\ of\ \frac{Insert}{Vector}$$

A 1:1 molar ratio of insert to vector was found to be successful for all purposes

Ligation reactions were set up as follows:

1 µl T4 DNA Ligase
100 ng Vector
x ng Insert
1 µl T4 Ligase Buffer (10X)
Nuclease free water to 10 µl

Reactions were incubated overnight at 4^{o}C, followed by transformation of 5 µl into competent DH5α cells. The cells were then streaked out on LB-Agar plates containing ampicillin and single colonies were selected and grown in LB media. After plasmid DNA isolation using Qiagen Mini-prep Kit, a PCR was carried out using the extracted plasmid DNA as a template with the same primer/reaction conditions as in the restriction digest step to check for the presence of the insert. This was then sequenced to verify that the insert was successfully inserted into the vector with no mutations.

2.2.10 Amplification of chromatin isolated by immunoprecipitation

The presence of certain genes isolated by protein-specific immunoprecipitation was detected by PCR amplification. Primers and annealing temperature used for each gene is detailed in Table 2.4

The PCR reaction mix using 2X Pfu Master Mix (Rovalab) was prepared as follows:

12.5 µl 2X Pfu Master Mix
5 µl Chromatin
0.13 µl Forward Primer (100 µM stock)
0.13 µl Reverse Primer (100 µM stock)
Nuclease-free water to 25 µl

Thermal cycling conditions:

95°C for 2 min

95°C for 20 sec

x °C for 40 sec

72°C for 2 min

Repeat steps 2-4 for 30 cycles

72°C for 5 min

Hold at 4°C

Table 2-4: Primers for validation of gene targets

Gene Name	***Forward***	***Reverse***	***Annealing Temp (°C)***
CHST11	GACATACGAGTGATCCCTACAC	GTCAGGGTTAAGGAAACTGC	55
CKS2	TCCACACAGTCGTCTAGG	CGAGGGAAGATAGGTACG	52
GBP3	GATCTAAGTCGTCTGAGTGC	TGTGTTGACTAGGGACATG	52
GRK5	AGCCCTCATCATCTACTTGG	GCCTGGAAAGCAAAGAAC	55
IKBKG	GGGTCCTTAGTGAAAAGCAG	CTGTGTCATGATGATGTTGG	55
IL12A	GTCCCTCAGCTACAAACAG	GGTCTAGAGATGAGCTTGG	52
KHDRBS1	CTCTCGAACTCCTAGCCTC	GGTTGACATTGTTGGTACG	54
KIAA0368	TCTCACATGTCCTATGTCC	GACACTCTAGTAATCAGCCG	52
NUB1	ACTCCTTCGTCCTTCTGTC	GACAAACGGCTTATCACAG	50
PRKAB2	CGCAGGGAAGAATTCAAG	CAGCTGTCAAAGCATTGC	56

RP11	GATGCGGAGAAACTTCAGAG	AGTTTGAGAACTGCTACTCTGG	55
SARS2	GAGGGCAATGTCACATAGC	GAACCTCCTGTACGAGTATGC	56
SDSL	ATCCGGGACAATCACTAGAC	CCTCTCATCAGCCTCTCATT	56
SLC22A13	CATGCAACCGTATTCAGC	CACCAAAGAACAGAGGTCC	55
TGS1	AGTCAGGCATGCATGTAGTC	CAGGCTAGTCTCCAACTTCC	56
WNT8A	CACACACAGGAAGTGAAAGC	GGTCTGCAAACTGACTGTTG	56
ZNF503	AGCCAGAAGAAGAAAAGAGG	GCCTCACTGCTACTCCTTC	54

Samples were run out on a 2% agarose gel at 70V for 1 h.

2.2.11 Site-directed mutagenesis

Site-directed mutagenesis for the generation of One-StrEP IRF-1 mutants was carried out as below.

Table 2-5: Primers used for site-directed mutagenesis

Gene to be cloned	***Vector backbone***	***Tag encoded by the vector***	***Primers (5' to 3')***	***DNA template***
IRF-1 Δenh	pEXPR-IBA-105	N-terminal One-STrEP	Forward: GCACTTCGGAATTCGCCCATCACTCGGATGCGC Reverse: GACGTATCGTCGGATCCCTACGGTGCACAGGGAATGGC	OneSTrEP-IRF-1 wt

IRF-1 Δ25	pEXPR-IBA-105	N-terminal One-STrEP	Forward: GCACTTCGGAATTCGCCCATCACT CGGATGCGC Reverse: GACGTATCGTCGGATCCCTAGTTC TTCAGATCTGTGAAGACACG	OneSTrEP-IRF-1 wt

12.5 µl 2X Pfu Master Mix

2.5 µl Band Doctor

50 ng Template DNA

0.13 µl Forward Primer (100 µM stock)

0.13 µl Reverse Primer (100 µM stock)

Nuclease-free water to 25 µl

Thermal cycling conditions:

95°C for 1 min

95°C for 50 sec

55°C for 1 min

68°C for 15 min

Repeat steps 2-4 for 18 cycles

68°C for 30 min

Hold at 4°C

After thermal cycling, 0.5 µl DpnI (10 U/ µl, Invitrogen) was added to each amplification product and incubated at 37^{o}C for 90 min. To inactivate the DpnI, the reaction mix was incubated at 65^{o}C for 10 min.

5 µl of the digested product was transformed into DH5α competent cells and streaked out on LB-Agar plates containing selective antibiotics. Single colonies were selected and used for growing starter cultures. Plasmid DNA was isolated

using the Qiagen Mini-prep Kit, as per the manufacturer's instructions and sequences prior to use.

2.3 General biochemical techniques

2.3.1 SDS-PAGE

Polyacrylamide gels were prepared using the recipes listed below using the Bio-Rad Protean II mini-gel system. The separating gel was poured first (about 70% of total gel) and overlaid with water to obtain an even surface and exclude oxygen from the gel. Once the separating gel had polymerized, the water was thoroughly removed and the stacking gel poured on top. A comb comprising 10 or 15 wells was immediately inserted and the gel allowed to polymerise. Once solidified, the gel was placed in a tank with 1X running buffer and the comb was removed. Prior to loading, samples were prepared by addition of either 2X sample buffer or 5X sample buffer and heated for 2-5 min at 95^{o}C. the mix was then loaded onto the gel along with pre-stained protein standard (Bio-Rad) and run at 120 V until the proteins exit the stacking gel and 150 V until the dye front reached the bottom of the gel.

2X Sample Buffer

300 mM Tris (pH6.8)
5% (w/v) SDS
25% (v/v) glycerol
200 mM DTT
bromophenolblue (as required)
aliquoted and stored at -20^{o}C

5X Sample Buffer

250 mM Tris (pH6.8)
12.5% (w/v) SDS, 125 mM
40% (v/v) glycerol
1 M DTT
bromophenolblue (as required)
aliquoted and stored at -20^{o}C

Separation Gel

8-15% acrylamide mix
0.39 M Tris-HCl (pH 8.8)
0.1% (w/v) SDS
0.1% (w/v) APS
0.04% (v/v) TEMED

Stacking Gel

5% acrylamide mix
0.13 M Tris-HCl (pH 6.8)
0.1% (w/v) SDS
0.1% (w/v) APS
0.04% (v/v) TEMED

1X Running Buffer

192 mM glycine
25 mM Tris-HCl (pH 6.8)
0.1% (w/v) SDS

2.3.2 Staining of SDS-PAGE gels

2.3.2.1 Coomassie staining

To visualize proteins by Coomassie brilliant blue staining, polyacrylamide gels were fixed for 10 min at room temperature and stained with coomassie blue stain for 30 min. Gels were then placed in destain until bands were of sufficient intensity (from 30 min to overnight) and dried onto chromatography paper using a heated vacuum gel dryer (Gel Master Model 1426, Welch Rietschle Thomas).

Fix

50% (v/v) methanol
10% (v/v) glacial acetic acid

Coomassie stain

50% (v/v) methanol
10% (v/v) glacial acetic acid
0.2% (w/v) coomassie blue R-250 (Sigma)

Destain

7.5% (v/v) methanol

10% (v/v) glacial acetic acid

2.3.2.2 Silver staining

Silver staining of SDS-PAGE gels was used to detect low concentrations of protein due to its high sensitivity. Silver staining was carried out using the Thermo Scientific Pierce Silver Stain Kit, as per the manufacturer's instructions. Briefly, polyacrylamide gels were washed 2 x 5 min in ultrapure water followed by fixation for 2 x 15 min in a mixture of ethanol and acetic acid. Gels were then incubated with 10% ethanol for 2 x 15 min, washed in ultrapure water and sensitized for 1 min using sensitizer solution. Following further washing, gels were stained using silver stain for 30 min, washed then developed with developer solution for 3-5 min. The developer solution was removed and the reaction was stopped with 5% acetic acid for 10 min.

2.3.3 Western Blotting

After separation by SDS-PAGE, proteins were transferred onto 0.2 µm nitrocellulose membranes (Whatman) using Bio-Rad transfer apparatus. The gel was placed in a cassette along with the membrane (pre-soaked in transfer buffer), filter paper and sponge. The transfer was carried out in tanks containing agitated transfer buffer and an ice pack to control the temperature for 1 h at 100 V. Once the transfer was complete, membranes were rinsed in phosphate-buffered saline (PBS) containing 0.1% Tween-20 (PBST) three times for 5 min each. To block any non-specific binding sites, the membrane was incubated in blocking solution [5% (w/v) dries skimmed milk powder (Marvel) in PBST] for 1 h at room temperature. The membrane was then incubated with primary antibody in blocking solution for 1 h at

room temperature or overnight at 4^{o}C, followed by washing as described above. Next, the membrane was incubated with the appropriate secondary antibody, diluted 1:2000 in blocking solution, for 1 h at room temperature then washed three times with PBST. To detect antibody signal, blots were incubated with a mixture of enhanced chemiluminescence (ECL) solutions I and II at a 1:1 ratio for 1 min, dried with tissue paper and exposed to Hyperfilm ECL (Amersham) or X-Ray film (SLS). The films were subsequently developed using a Konica Medical Film Processor (SRX-101A).

1X Transfer Buffer

192 mM glycine
25 mM Tris
20% (v/v) methanol

10X PBS

1.37 M NaCl
0.1 M Na_2HPO_4
27 mM KCl
18 mM KH_2PO_4

ECL Solution I

100 mM Tris-HCl (pH 8.5)
2.5 mM Luminol
0.4 mM p-Coumaric acid

ECL Solution II

100 mM Tris-HCl (pH 8.5)
0.02% (v/v) H_2O_2

The primary antibodies used in this thesis are listed below in Table 2-6.

Table 2-6: Primary Antibodies

Target	***Type***	***Name***	***Supplier***	***Dilution***
GAPDH	Mouse Monoclonal	9489	Abcam	1:3000
Histone H3	Mouse Monoclonal	-	Abcam	1:1000
IgG				
IRF-1	Mouse Monoclonal	20/IRF-1	BD	1:1000
IRF-1	Rabbit Polyclonal	H-205	Santa Cruz	1:1000

Kap-1	Rabbit Polyclonal	-	Bethyl	1:1000
MDM2	Mouse Monoclonal	2A10	Moravian Biotechnology	1:1000
MDM2	Mouse Monoclonal	4B2	Moravian Biotechnology	1:1000
ZNF350	Rabbit Polyclonal	77085	Abcam	1:1000
ZNF350	Rabbit Polyclonal		Sigma	1:1000

2.3.4 Stripping nitrocellulose blots

To strip antibodies from nitrocellulose membranes to enable the addition of subsequent antibodies, the membranes were washed in water three times then incubated with 0.2 M NaOH for 10 min at room temperature with gentle agitation. Blots were then rinsed in water three times then incubated at -20^{o}C for approximately 20 min. Following further washing in water, blots were then blocked in 5% milk/PBST for 1 h and incubated with further antibodies as required.

2.4 Cell culture

All tissue culture disposables such as culture plates, flasks and pipettes were from TPP or Grenier-Cellstar unless otherwise stated.

2.4.1 Cell lines

Table 2-7 lists the cell lines used in this thesis along with their origins, optimum growth conditions and culture media. A375 cells are derived from human malignant

melanoma and were chosen for most studies in this thesis, unless otherwise stated, due to the high basal expression levels of both ZNF350 and IRF-1.

Table 2-7: Cell lines

Cell line	*Origin*	*Growth conditions*	*Culture media*
A375	Human (malignant melanoma)	10% CO_2, 37^oC	DMEM
H1299	Human (non-small cell lung carcinoma)	5% CO_2, 37^oC	RPMI-1640
HeLa	Human (cervical cancer)	10% CO_2, 37^oC	DMEM
U251	Human (glioblastoma)	5% CO_2, 37^oC	DMEM
SF539	Human (glioblastoma)	5% CO_2, 37^oC	DMEM

2.4.2 Sub-culturing of cells

Cells were grown in 10 cm diameter plates and sub-cultured when cells reach approximately 100% confluency (2-3 times a week depending on cell type). Briefly, the medium was discarded and adherent cells were rinsed once with sterile PBS, followed by incubation at 37^oC for 5 min with 2 ml Trypsin-EDTA (Invitrogen). Once the cell monolayer had detached from the plate, 8 ml of fresh media was added to deactivate the trypsin-EDTA, and the cells were mixed by gentle pipetting. 1 ml of this cell suspension was then added to a new culture dish containing 9 ml of pre-warmed fresh media, giving a 1:10 dilution.

2.4.3 Freezing and thawing cells

To freeze for storage, cells were grown to approximately 80-100% confluency and trypsinised as above. The trypsinsied cells were collected by centrifugation (1000 rpm for 5 min at room temperature) and the cell pellet resuspended in 3 ml of freezing media. The cells were then transferred to cryotubes (Nunc) in 1 ml aliquots and frozen gradually in a Nalgene cryo-freezing container at -80^{o}C overnight. The frozen cells were then transferred to liquid nitrogen for long-term storage.

To recover the frozen cells, they were warmed to 37^{o}C and transferred to a sterile culture dish containing fresh pre-warmed media followed by overnight incubation to allow the cells to adhere to the plate. The next day, the media was replaced to remove any residual DMSO that was present in the freezing media.

Freezing media

40% (v/v) culture media
50% (v/v) FBS
10% (v/v) DMSO

2.4.4 Transient transfection of DNA

Cells were transfected at approximately 70-80% confluency using Attractene (Qiagen) according to the manufacturer's instructions. DNA levels between samples were normalised using the corresponding empty vector and cells were harvested 16-24 h post-transfection.

2.4.5 Drug treatments

Several experiments described in this thesis involved the treatment of cells with various drugs prior to harvest and lysis. Table 2-7 summarises the drugs used, the concentration and duration of treatment and the main effects of each drug. Details of specific use are indicated in the relevant sections.

Table 2-8: Drug Treatments

Drug	*Function*	*Concentration*	*Time*
DRB	Transcription inhibitor	5-100 µM (as indicated)	24 h
Etoposide	DNA damage mimetic	5-100 µM (as indicated)	4 h
IFNγ	Cytokine	50-1000 U/ml (as indicated)	4 h
Poly(I:C)	Viral dsRNA mimetic	50 µg/ml	4 h
Trichostatin A	Histone deacetylase inhibitor	1 µM	24 h

2.4.6 Cell irradiation

Cells were irradiated in tissue culture dishes after removal of the lid, with 5-100 j/m^2 of UV light using a UV crosslinker as per the manufacturer's instructions.

2.4.7 In vivo crosslinking

2.4.7.1 DSP (dithiobis[succinimidylpropionate])

Mammalian cells in 10 cm plates were washed twice with PBS then incubated with 4.5 ml PBS and 20 mM DSP (Thermo; final concentration 2 mM) diluted in 500 µl DMSO for 30 min at room temperature. To stop the reaction, 100 µl of 1 M Tris pH 7.5 (final concentration 20 mM) was added to the plates and incubated for 15 min at room temperature. Cells were then harvested as below and lysed in lysis buffer, omitting DTT.

2.4.7.2 Formaldehyde

Formaldehyde (37%) was added directly to cell media to a final concentration of 1% and incubated for 10 min at room temperature with gentle rocking. The reaction was stopped by the addition of glycine at a final concentration of 0.125 M and incubated for 5 min at room temperature with gentle rocking. Cells were then harvested as below and lysed in lysis buffer, omitting DTT.

2.4.8 Cell harvesting and lysis

For harvesting, cells were placed on ice and the media discarded. The cells were rinsed with ice-cold PBS (1 ml/well for a 6 well plate or 10 ml for a 10 cm plate) and then harvested directly into lysis buffer (200 µl/well for a 6 well plate or 1 ml per 10 cm plate). Samples were vortexed briefly, incubated on ice for 20 min and then centrifuged at 16000 g for 15 min at 4^{o}C. The supernatant (lysate) was transferred to a fresh tube and either used immediately in further assays or snap-frozen in liquid nitrogen and stored at -80^{o}C.

0.2% Triton Lysis Buffer

50 mM Hepes (pH 8.0)
150 mM NaCl
10 mM NaF
0.2 % (v/v) Triton X-100
2 mM DTT
0.1 mM EDTA
1X protease inhibitor mix (see below)

Protease Inhibitor Mix (10X stock)

200 µg/ml leupeptin
10 µg/ml aprotinin
20 µg/ml pepstatin
10 mM benzamidine

100 µg/ml soybean trypsin inhibitor
20 mM pefabloc
10 mM EDTA

2.5 Protein expression and purification from *E.coli*

2.5.1 Expression of His-tagged ZNF350

pCOLD-ZNF350 bacterial plasmid was transformed into *BL21-DE3* or *BL21-DE3** cells as required and streaked onto LB-Agar plates. A single colony was selected from the LB-Agar plate, inoculated into 50 ml LB containing ampicillin (50 µg/ml) and incubated overnight in at 37^{o}C with shaking (220 rpm). The following day, the 50 ml culture was diluted in 1 litre of fresh LB + ampicillin (50 µg/ml) and incubated at 37^{o}C until the OD_{600} reached 0.4. The culture was rested at 15^{o}C for 30 min following which protein expression was induced by the addition of 0.5 mM IPTG and incubated for 15 – 240 min at 15^{o}C in a shaker incubator. Cells were then collected by centrifugation at 6000 g for 20 min at 4^{o}C and the supernatant discarded. The cell pellet was snap-frozen in liquid nitrogen and when required, the cells were resuspended in the required lysis buffer (10-15 ml per pellet collected from 500 ml culture) and incubated for 30 min on ice. The cells were again snap-frozen in liquid nitrogen and rapidly thawed using a water bath set at room temperature. This freeze-thaw cycle was repeated, after which the cells were briefly sonicated using a Soniprep 150 tip sonicator (Sanyo) for 3 x 15 sec at amplitude 10 microns with 15 sec incubations on ice in between. The sonicated cells were centrifuged at 16000 g for 15 min at 4^{o}C and the lysate (supernatant) was collected.

Lysis Buffer 1	*Lysis Buffer 2*
20 mM Tris (pH 8)	25 mM Tris (pH 8)
300 mM NaCl	300 mM NaCl
0.5 % NP-40	1x protease inhibitor mix (-EDTA)

20mM imidazole

1 mg/ml lysozyme

1x protease inhibitor mix (-EDTA)

Lysis Buffer 3

50 mM Na_2HPO_4 (pH 8)

300 mM NaCl

8 M Urea

1x protease inhibitor mix (-EDTA)

Lysis Buffer 4

50 mM Na_2HPO_4 (pH 8)

300 mM NaCl

1x protease inhibitor mix (-EDTA)

2.5.2 Purification of His-tagged ZNF350

E.coli lysate (containing no DTT) from a 500 ml culture was mixed with 1 ml (50 % slurry) Ni^{2+}-NTA agarose beads (Qiagen; pre-washed 2X in Wash Buffer I) and incubated for 1 h at 4^{o}C on a rotating table. The mix was transferred to a 10 ml disposable column (MoBiTec) and allowed to empty by gravity flow. The beads were then washed 2X with 5 ml Wash Buffer I, 2X with 5 ml Wash Buffer I+ATP and 3X with 5 ml Wash Buffer II. Subsequently, 5 ml of Elution Buffer was added to the beads, incubated for 30 min at 4^{o}C on a rotating table and the eluate collected by gravity flow. The eluate was then exchanged into Protein Buffer I using a HiTrap desalting column on an AKTA chromatography system and stored at -80^{o}C.

Wash Buffer I

20 mM Tris (pH8)

300 mM NaCl

0.5% NP-40

20mM imidazole

2mM benzmidine

Wash Buffer I + ATP

As for Wash Buffer I but with:

10 mM $MgCl_2$

5mM ATP

Wash Buffer II

As for Wash Buffer I but with:

40 mM imidazole

Elution Buffer

20 mM Tris (pH 8)

150 mM NaCl

300 mM imidazole

Protein Buffer I

20 mM Tris (pH8)

150 mM NaCl

2.6 Assays

2.6.1 Peptide affinity chromatography (peptide pulldown)

A375 cells were grown in 10 dishes as required and lysed in 1 ml 0.1% Triton (v/v) lysis buffer as described in section 2.4.8. The cell lysate was collected and treated and pre-cleared using sepharose-4B beads (Sigma) for 1 h at 4°C with gentke shaking. 200 µl beads (50% slurry) were used per 10 cm plate of cells and washed 3 times in PBS prior to use. The beads were collected by centrifugation at 500 g for 3 min at 4°C and the supernatant collected and protein levels quantified using Bradford's assay.

Concurrently, peptide affinity columns were prepared using Mobicol columns (MoBiTech) containing 35 μm pore-size filters and a Luer-lock cap then adding 100 μl of streptavidin-agarose beads (Invitrogen, 50% slurry) which were washed 3 times with 600 μl PBS. Post-washing, the bottom of the column was sealed using the Manufacturer-supplied plug and enough biotinylated peptide to saturate the streptavidin-agarose (approximately 2 μl of 5 mg/ml peptide stock diluted in 200 μl PBS) was added and incubated with the beads for 1 h at room temperature on a rotating table. The plug on the column was removed to allow the column to empty by gravity flow and the beads were then washed 3 times with 600 μl PBS (or Buffer W) to remove any unbound protein.

The column was sealed again with the plug and pre-cleared lysate (0.5 mg) prepared as described above, was added to the peptide column and incubated with the peptide-bound bead mixture for 1 h at room temperature on a rotating table. The outlet plug on the column was once again removed to allow the column to empty by gravity flow, and then washed 4 times with 600 μl PBS supplemented with 0.2% Triton X-100 (v/v) followed by 2 washes with 600 μl PBS alone (or Buffer W). The resin was then emptied into a microfuge tube by adding 200 μl PBS (or Buffer W) to the column and dislodging the filters using micropipette tips. The tubes were then centrifuged to collect the resin and the supernatant was discarded. Next, 50 μl SDS sample buffer supplemented with 100 mM DTT was added to the resin and boiled at 85°C for 5 min to elute the bound proteins from the peptide-bead complex. The beads were then pelleted by centrifuging at 500 g for 3 min at room temperature and the eluate collected and transferred to a new tube for storage at -20°C. eluates were run out on a 10% SDS-PAGE gel and transferred to nitrocellulaose and immunoblotted as required.

Buffer W

100 mM Tris (pH 8.0)
150 mM NaCl
1 mM EDTA

2.6.2 Protein binding assays

2.6.2.1 Direct protein binding assay

A white 96-well microtitre plate (Fisher) was coated with 100 ng of protein 1 (His-ZNF350) in 50 µl 0.1 M $NaHCO_3$ buffer (pH 8.6) and incubated overnight at 4^oC. E following day the plate was washed three times with 200 µl PBST and non-reactive sites were blocked using 200 µl PBS containing 3% BSA (w/v) for 1 h at room temperature. The plate was then washed as previously and a titration of protein 2 (GST-IRF-1, kindly gifted from Vivien Landré and Carrie-Anne Sharma) in 50 µl of 1X ELISA reaction buffer was added to the coated wells and incubated for 1 h at room temperature. The plate was washed again three times with 200 µl PBST and incubated with primary antibody diluted in 50 µl of PBS+BSA for 1 h at room temperature. The wells were washed again as above and incubated with HRP-conjugated secondary antibody diluted in 50 µl of PBS+BSA for 1 h at room temperature. After a final set of 3X PBST washes, binding was detected by electrochemical luminescence (50 µl ECL mix per well for 1 min incubation) and quantified using a luminometer (Labsystems; Fluoroskan Ascent FL).

1X ELISA Reaction Buffer

25mM HEPES (pH 7.5)
50 mM KCl
10 mM MgCl2
5% (v/v) glycerol
0.1% (v/v) Tween-20
2 mg/ml BSA

2.6.2.2 Direct peptide binding assay

A white 96-well microtitre plate (Fisher) was coated with 100 ng of His-ZNF350 in 50 µl 0.1 M $NaHCO_3$ buffer (pH 8.6) and incubated overnight at 4^oC. The following day, the wells were washed three times with 200 µl PBST and non-reactive sites

were blocked using ELISA blocking buffer (Thermo Scientific) or 200 µl PBS containing 3% BSA (w/v) for 1 h at room temperature. After another wash step, saturating concentrations of biotin tagged peptides (~60 pmol; 0.25 µl of 5 mg/ml stock) in 50 µl PBS were added to the plate and incubated for 1 h at room temperature. Peptides used for direct peptide binding assays were synthesised with an N-terminal Biotin tag and Ser-Gly-Ser-Gly spacer and sourced from Chiron Mimotopes. The wells were washed three times as above and then incubated with HRP-conjugated secondary antibody diluted in 50 µl PBS+BSA. After a final wash step, binding was detected by electrochemical luminescence (50 µl ECL mix per well for 1 min incubation) and quantified using a luminometer (Labsystems; Fluoroskan Ascent FL).

Table 2-9: IRF-1 Peptides

Number	***Residues***	***Sequence***
1	1-20	MPITRMRMRPWLEMQINSNQ
2	16-35	INSNQIPGLIWINKEEMIFQ
3	31-50	EMIFQIPWKHAAKHGWDINK
4	46-65	WDINKDACLFRSWAIHTGRY
5	61-80	HTGRYKAGEKEPDPKTWKAN
6	76-95	TWKANFRCAMNSLPDIEEVK
7	91-110	IEEVKDQSRNKGSSAVRVYR
8	106-125	VRVYRMLPPLTKNQRKERKS
9	121-140	KERKSKSSRDAKSKAKRKSC
10	136-155	KRKSCGDSSPDTFSDGLSSS
11	151-170	GLSSSTLPDDHSSYTVPGYM
12	166-185	VPGYMQDLEVEQALTPALSP
13	181-200	PALSPCAVSSTLPDWHIPVE
14	196-215	HIPVEVVPDSTSDLYNFQVS
15	211-230	NFQVSPMPSTSEATTDEDEE
16	226-245	DEDEEGKLPEDIMKLLEQSE
17	241-260	LEQSEWQPTNVDGKGYLLNE
18	256-275	YLLNEPGVQPTSVYGDFSCK
19	271-290	DFSCKEEPEIDSPGGDIGLS
20	286-305	DIGLSLQRVFTDLKNMDATW
21	301-320	MDATWLDSLLTPVRLPSIQA
22	306-325	LDSLLTPVRLPSIQAIPCAP

2.6.3 Immunoprecipitation (IP)

2.6.3.1 OneSTrEP pulldown

A375 cells were grown in 10 cm plates and transfected with 4 µg OneSTrEP-IRF-1 wt, OneSTrEP-IRF-1 Δenh, OneSTrEP-IRF-1 ΔC25, OneSTrEP empty vector. 24 h post transfection, cells were scraped into 1 ml 0.2% Triton lysis buffer as described in 2.4.8. Lysate (5 mg) was added to 40 µl (50% slurry) streptactin macroprep (IBA), which had been pre-washed three times in Buffer W, and incubated for 1 h at 4^{o}C on a rotating table. The macroprep resin was then pelleted by centrifugation at 1000 g for 5 min at 4^{o}C and washed five times with 0.5 ml Buffer W. For each wash, samples were gently mixed by inversion and then centrifuged as above to pellet the resin, after which the supernatant was discarded. Following the final wash, OneSTrEP-proteins were eluted by the addition of 100 µl Buffer E for 30 min at 4^{o}C on a rotating table. Samples were centrifuged as above and the supernatant containing OneSTrEP-proteins was collected. Samples were loaded onto a polyacrylamide gel and analysed by SDS-PAGE/immunoblot.

Buffer W

100 mM Tris (pH 8.0)
150 mM NaCl
1 mM EDTA
1 mM benzamidine

Buffer E

2 mM biotin in Buffer W

2.6.3.2 Endogenous Protein IP

2.6.3.2.1 Cross-linking antibodies to protein G sepharose beads

6.8 µg of either anti-ZNF350 (Abcam) or anti-IRF-1 (BD) antibody was mixed with 100 µl packed protein G sepharose beads (Amersham GE) in 1 ml PBS and incubated at 4°C for 1 h on a rotating table. The beads were washed twice with 10 volumes of 0.2M sodium borate (pH 9) by centrifugation at 3000 g for 5 min to pellet the beads and the supernatant discarded. The beads were then resuspended in 10 volumes of 0.2M sodium borate (pH 9) and to check for cross-linking efficiency, a 10 µl aliquot of the beads ('before') was collected at this stage to use as a control. To the remainder of the beads, dimethyl pimelimidate (DMP; Pierce; prepared fresh) was added to the beads at a final concentration of 20 mM, and incubated for 2 h at room temperature with gentle mixing. Samples were centrifuged as above and the supernatant discarded. To stop the reaction, the beads were washed once on 0.2 M ethanolamine (pH 8) for 2 h at room temperature. The samples were centrifuged once again as above and the beads were resuspended in PBS to make a 50% slurry. A further 10 µl aliquot of the beads ('after') was collected to check for cross-linking efficiency and the remainder of the beads stored at 4°C until use.

The 'before' and 'after' aliquots were boiled in 2X sample buffer for 5 min and then run on an SDS-PAGE gel. Good antibody-bead coupling is signified by the presence of heavy chain (55 kDa) bands in the 'before' but not the 'after' lanes, when observed by coomassie staining.

2.6.3.2.2 Immunopreciptation using antibodies coupled to protein G-sepharose beads

A375 cells were grown in 10 cm plates as required and lysed in IP Lysis Buffer at approximately 95% confluency. Lysate (2 mg) was incubated with cross-linked beads (10 µl packed volume) for 2 h 30 min at 4°C on a rotating table. Beads were pelleted by centrifuging at 500 g for 2 min at 4°C and the supernatant discarded. The beads were then washed as follows: two times with 1 ml wash buffer supplemented with 0.5 M NaCl then two times with 1 ml wash buffer. Beads were centrifuged as

above for each wash then the supernatant discarded. After the final wash, bound proteins were eluted by heating in 50 µl 2X sample buffer at 95°C for 5 min. Beads were collected b centrifuging at 500 g for 3 min at room temperature and the superatant transferred to a new eppendorf tube. 25 µl of sample was then loaded on to a 5-10 % polyacrylamide gel for analysis by SDS-PAGE/immunoblot as required.

IP Lysis Buffer

20 mM Tris (pH 7.4)
150 mM NaCl
1 mM EDTA
1 mM DTT
1 % NP-40
10% glycerol
1 mM Na_3VO_4
1X PIM

Wash Buffer

25 mM HEPES (pH 7.5)
10% glycerol
1 mM EDTA
0.1% NP-40
50 mM NaF
0.5 mM Na_3VO_4
0.2 mg/ml benzmidine

2.6.4 Dual luciferase reporter assays

A375, H1299 or HeLa cells were cultured in 24-well plates and transfected with pCMV-RenillaLuc (1-2 ng per well) along with pLTR-Luc-HIV-1 (1 ng per well), MMP9-Luc (100 ng per well) and 1.5 µl Attractene per well. For experiments where the over-expression of IRF-1 and/or ZNF350 on promoter activity was studied, a

titration of plasmids encoding these proteins (as indicated in the figure legends) was co-transfected with the pCMV-RenillaLuc and relevant luciferase plasmids.

Luciferase assays were performed 24 h post-transfection, or at time points indicated in figure legends, using the Dual Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions. Briefly, cells were washed in ice-cold PBS and lysed by incubating with 100 µl passive lysis buffer (supplied by manufacturer) per well for 15 min at room temperature with shaking. The lysed cell suspension was mixed by pipetting and transferred to microfuge tubes. Samples were centrifuged at 16000 g for 5 min at 4°C to collect the cell debris and the lysate was transferred to a fresh tube. Cell lysate (5 µl per well) was added to a white 96-well microtitre plate (Fisher) and mixed with 25 µl luciferase assay reagent (LAR II; supplied by manufacturer). Immediately, firefly luciferase luminescence was measured using a Fluroskan Ascent F1 luminometer (Labsystems). Next, 25 µl Stop and Glo reagent (supplied by manufacturer) was added to each well and the renilla luciferase signal quantified using a luminometer. Signals were normalised using renilla luciferase signal as an internal control and expressed as a ration of firefly : renilla luciferase. Results are represented as mean ± SD.

2.6.5 Proximity Ligation Assay

A375 cells were seeded in 6-well plates with coverslips placed in the bottom of each well. If required, the cells were treated as follows: 20 µM DRB for 24 h, 400nM TSA for 24 h, 20 µM etoposide for 4 h, 200 U/µl IFNγ for 4 h or 50 µg/ml poly(I:C) for 4 h.

The cells were washed 3 times for 5 min in 1 ml PBS with gentle shaking then fixed in 1 ml Fixation buffer for 15 min at room temperature with gentle shaking. After a further 3 washes, as described above, the cells were permeablised in 1 ml of 1% Triton X-100 diluted in PBS for 10 min at room temperature. Cells were washed a further 3 times then blocked with 1 ml of 3% BSA in PBS for 30 min at room temperature. The blocking buffer was removed and the cell-containing coverslips

were transferred to a new dish. Due to the PLA solution being sensitive to light, the dish was covered in foil and the cells kept in the dark as much as possible. Anti-ZNF350 (Abcam 77085) and/or anti IRF-1 (BD) antibodies diluted in 3% BSA/PBS were then added to the cells and incubated at 4°C overnight. A damp tissue was placed in the dish to ensure the cells did not dry out.

The next day, the coverslips were washed 3 times in PBS for 5 min at room temperature and 20 µl PLA Probe Mix was added and incubated at 37°C for 1h. Following 3 washes for 5 min in Wash Buffer A, 20 µl of Ligation Mix was added to the cells and incubated at 37°C for 30 mins. Cells were again washed twice in Wash Buffer A for 5 min then 20 µl of Polymerase Amplification Mix was added and incubated at 37°C for 2h. A further 2x 5 min washes in Wash Buffer B was followed by a final 5 min was with 1% Wash Buffer B diluted in water.

To mount the cells on a slide, 1 µl DAPI was mixed in a few drops of fluorescent mounting medium (Dako) and approximately 5 µl was added to a clean slide. The coverslip was placed cell-side down on top of the DAPI medium and pressed gently to remove any air bubbles. Excess liquid was removed by gentle blotting and the edges of the coverslip was sealed using clear nail polish. The slides were covered in foil and allowed to dry at 4^{o}C. Cells were viewed using an Olympus BX51 microscope and fluorescence measured and analysed using ImageJ.

Fixation Buffer

3.7% Formaldehyde
10 mM EGTA pH 8.0
100 mM PIPES pH 6.8
1 mM MgCl
0.2% Triton X-100

Wash Buffer A

100 mM Tris (pH 7.4)
0.15 M NaCl

0.05% Tween 20

Wash Buffer B

200 mM Tris (pH 7.4)
0.1 M NaCl

PLA Probe Mix, Ligation Mix and Polymerase Mix (Duolink) were made up as per the manufacturer's instructions

2.6.6 Sucrose gradient fractionation

A375 cells were grown in 15 cm dishes then harvested by addition of 2 ml Trypsin-EDTA to dissociate the cells from the plate surface. This was collected by addition of 10 ml of PBS containing 10% FBS and 20 mM DTT. The cells were pelleted by centrifugation at 1400 rpm for 4 min at 4^{o}C and re-suspended in 3 ml NBA. 3 ml of NBB was added and cells were incubated for 3 min on ice and pelleted by centrifuging for 4 min at 2000 rpm at 4^{o}C. The supernatant was discarded and the nuclei washed in 10 ml NBR followed by re-suspension in 500 µl NBR. To prepare soluble chromatin, nuclei was digested by incubation with 10 units of micrococcal nuclease and 5 µl per ml RNaseA/T1 (Ambion) for 10 min at room temperature. The reaction was stopped by addition of EDTA at a final concentration of 10 mM. The samples were collected by centrifuging for 30 sec at 5000 rpm, re-suspended in 500 µl Teep80 then incubated on ice overnight to allow for cells lysis. To remove cell debris, samples were centrifuged for 5 min at 13 000 rpm at 4°C and soluble chromatin was fractionated using 10-50% (w/v) isokinetic sucrose gradients in TEEP80 by centrifugation for 120 minutes at 50 000 rpm in a MLS 50 Beckman rotor at 4°C. Gradients were fractionated in 500 µl fractions by upward displacement and the absorbance profile was monitored continuously to ensure proper fractionation. Each fraction was aliquoted into two; one half was used to analyse the protein content by ethanol precipitation, followed by washing to remove all sucrose

and then re-suspended in SDS-sample buffer and analysed by SDS-PAGE/immunoblot. The other half was used to extract the DNA with phenol/chloroform and concentrated by ethanol precipitation. DNA pellets were washed, re-suspended in H_2O and analysed on a 1% agarose gel.

NBA

10 mM Tris-HCl (pH 7.6)
85 mM KCl
5.5% (w/v) sucrose
0.5 mM spermidine
0.2 mM EDTA
0.05 mM Pefabloc
20 mM DTT

NBB

NBA + 0.1% (v/v) IGEPAL

NBR

10 mM Tris-HCl (pH 7.6)
85 mM KCl
5.5% (w/v) sucrose
1 mM $CaCl_2$
1 mM $MgCl_2$
0.1 mM pefabloc

Teep 80

10 mM Tris-HCl (pH 7.6)
80 mM NaCl
1 mM EDTA
1 mM EGTA
0.1 mM pefabloc

2.7 Chromatin Immunoprecipitation

2.7.1 Crosslinking

A375 cells were crosslinked by adding formaldehyde directly to the media at a final concentration of 1%. The cells were then incubated for 10 min at room temperature with gentle rocking. To stop the cross-linking reaction, glycine was added to the plate at a final concentration of 0.125M and incubated for 5 min at room temperature

with gentle rocking. Cells were collected by centrifuging at 500 g for 5 min and the pellet washed twice in PBS. The cell pellet can be snap-frozen and stored at -20^{o}C.

2.7.2 Lysis

For ChIP experiment, A375 cells are lysed in a two-step process. The cell pellet is resuspended in Cell Lysis Buffer (7.5 ml per 15 cm plate) and incubated on ice for 15 min. Cells were collected by centrifuging at 16000 g for 5 min at 4^{o}C then resuspended in Nuclei Lysis Buffer (1 ml to 2.5 ml per 15 cm plate, or ~10^{8} cells).

Cell Lysis Buffer

5 mM Hepes pH 8.0
85 mM KCl
0.5% NP-40
1X protease inhibitor mix (PIM)

Nuclei Lysis Buffers:

NLB1

50 mM Hepes pH 7.5
140 mM NaCl
1 mM EDTA
0.5% (v/v) NP-40
0.25% (v/v) Triton X-100
1X PIM

NLB2

50 mM Hepes pH 8.0
250 mM NaCl
50 mM NaF
5 mM EDTA
1% (v/v) NP-40
1X PIM

NLB3 (RIPA Buffer)

50 mM Tris-HCl pH 8.0
150 mM NaCl
2 mM EDTA
1% (v/v) NP-40

NLB4

10 mM Hepes pH 7.5
2 mM MgCl2
10 mM NaF
10 mM KCl

0.5% (v/v) Sodium deoxycholate
0.1% or 1% (w/v) SDS
1X PIM

1 mM EDTA
1 mM EGTA
1 mM DTT
1X PIM

2.7.3 Chromatin shearing

The nuclear lysate was sheared using the either a (a) Bioruptor (Diagenode UCD-200) on the HIGH setting for specified number of cycles with 30 seconds on and off or (b) Soniprep 150 sonicator (Sanyo) at amplitude 10 for specified number of pulses with 1 min incubation on ice in between. The lysate was collected by centrifugation at 16000 g for 15 min at 4^{o}C and diluted 10-fold using IP Dilution Buffer. To check sonication efficiency, a 20 µl aliquot was taken and 1 µl of Proteinase K was added and the mixture incubated at 95^{o}C for 5 min. These samples were then run on a 1% agarose gel to examine the size of the DNA fragments. The rest of the samples were snap-frozen and stored at -80^{o}C.

IP Dilution Buffer

16.7 mM Tris pH 8.1
167 mM NaCl
1.2 mM EDTA
0.01% (w/v) SDS
1.1% (v/v) Triton X-100
1X PIM

2.7.4 Immunoprecipitation

500 µl of chromatin preparation was set aside for use as an Input control. Anti-ZNF350 (Sigma or Abcam 77085, as specified), anti-IRF-1 (Santa Cruz or BD, as specified) or anti-IgG (NEB 2729) antibodies were coupled to protein G sepharose beads as described in section 2.6.3.2.1. 2 ml of chromatin preparation was added to 80 µl antibody-coupled beads and incubated overnight at 4°C on a rotating table. The

next day, beads were pelleted by centrifuging at 500 g for 2 min at 4°C and the supernatant discarded. The beads were washed once with 1 ml Low Salt Wash Buffer, once with 1 ml High Salt Wash Buffer and once with 1 ml LiCl Wash Buffer. Beads were incubated for 3 min per wash on a rotating table and collected by centrifugation at 500 g for 2 min at 4°C. The protein-DNA bound beads were transferred to a new tube to remove DNA that is stuck to the sides and washed in 1ml TE Buffer with mixing for 1 min on a rotating table then collected by centrifuging as above. All the wash steps were performed in a cold room at 4°C. The bead pellet was then resuspended in 250 µl IP Elution Buffer, vortexed to mix and incubated for 1h at 65°C with mixing every 10 min. The eluate was collected by centrifuging at 500 g for 2 min at room temperature and transferred to a new microfuge tube. The elution step was repeated to collect remaining DNA and eluates pooled.

IP Elution Buffer

100 mM $NaHCO_3$
1% (w/v) SDS

2.7.5 Reversal of cross-links and DNA purification

20 µl of 5M NaCl was added to the Input and eluted samples and incubated at 65^{o}C with shaking overnight. The next day, 2.5 µl RNase cocktail (Invitrogen) was added to samples and incubated for 1 h at 37^{o}C followed by the addition of 20µl of 0.25 M EDTA, 20 µl of 1 M Tris-HCl pH 6.5 and 5 µl of 20 mg/ml Proteinase K and incubation for 2 h at 45^{o}C. The DNA was purified using the Qiagen PCR Purificiation Kit and the DNA eluted from the with 30-50 µl nuclease-free water. The chromatin was quantified using a Nanodrop ND-1000 spectrophotometer against nuclease-free water blank. The DNA samples are stored at -80^{o}C.

2.7.6 High-throughput sequencing of ChIP samples

ZNF350- and IRF-1-ChIP samples with biological replicates and Input controls (not immunoprecipitated) as prepared as above, were sent to The Genepool sequencing facility where they performed library preparation and sequencing using the Illumina HiSeq 2000, as described [156] and per the manufacturers instructions.

2.7.7 Analysis of ChIP-seq data

Sequencing data was analysed using the web-based tool, Galaxy [157] according to protocols provided by several online tutorials [158; 159]. Briefly, raw sequencing data was uploaded to Galaxy and quality statistics were generated using the FASTQC tool then the reads mapped to the hg19 reference genome using BWA. Peaks were called using MACS with the tag length set to 50 and MFOLD to 10. Biological replicates were intersected using the "Operate on genomic intervals" tool and common peaks for each condition were identified. Peaks were annotated by importing genomic data from the UCSC genome browser to Galaxy then intersecting the two data sets.

2.8 Peptide Phage display

Peptide phage display was carried out using the Ph.D Phage Display 12-mer library (NEB) as per the manufacturer's instructions. Some changes were made to the recommended procedures and are described below.

2.8.1 Panning Procedure

A white 96-well microtitre plate (Costar) was coated with recombinant His-ZNF350 diluted in 150 µl coating buffer (0 .1 M NaHCO3, pH 8 .6) at a concentration of 1 µg/ml per well. The plate was incubated at 4°C overnight with gentle agitation in a humidified container. The next day, coating solution was removed and blocking buffer (3% w/v BSA) was added and incubated for at least 1 h at room temperature with gentle agitation. The wells were washed in one of two ways: wash buffer added then removed rapidly 6 times with TBST (TBS + 0.1% [v/v] Tween-20) or wash buffer (TBST) added to the plate and incubated for 5 min then removed and repeated a further 5 times. The peptide library was added to wells at a concentration of 10^{11} phages diluted in TBST and incubated for 1 h at room temperature with gentle agitation. After a further round of washing, either fast or slow as described above, the phage particles were eluted with 100 µl of 0.2 M glycine-HCl (pH 2.2), 1 mg/ml BSA with gentle rocking for 10 min. This was then neutralized with 15 µl of 1M Tris-HCl pH 9.1 and eluates store at 4°C.

Eluted phages were amplified by infection of ER2378 cells and phage particles precipitated (described below). Biopanning was repeated and approximately 10^{11} pfu of the first round of amplified eluate was used as input phage for the second round.

2.8.2 Phage amplification

LB-Agar plates containing tetracycline (20 µg/ml) were prepared and streaked with the ER2738 *E. coli* host strain. Plates were incubated at 37°C overnight with inversion. A colony was picked from the plate and added to 20 ml LB which was then incubated at 37°C with shaking until OD_{600nm} reached 0.05. The phage eluates from above (70%) were added to the early-log phase ER2738 culture and inchubated for a further 4.5 h at 37°C with agitation. The supernatant was collected by centrifuging at 12,000 g for 10 min at 4°C, transferred to a new tube then re-centrifuged. The upper 80% of the supernatant was transferred to a fresh tube and 1/6 volume of 20% PEG/2 .5 M NaCl was added and the phage allowed to

precipitate at 4°C overnight. The following day the precipitated phages were spun at 12,000 g for 15 min at 4°C, the supernatant discarded and the sample was centrifuged once more to remove any residual supernatant. The pellet was resuspended in 1 ml 1X TBS, transferred to new tube and centrifuged at 16,000 g for 5 min at 4°C to pellet any residual cells. The supernatant was then transferred to another fresh tube and precipitated with 1/6 volume of 20% PEG/2 .5 M NaCl and incubated on ice for 1 h. Samples were centrifuged at 16,000 g for 5 min at 4°C to pellet the phages and the supernatant discarded, followed by a further re-centrifuging to remove any residual liquid. The amplified phage pellet was resuspended in 200 µl 1X TBS and after incubation at 65°C for 15 min to kill residual bacteria was stored at 4°C for a maximum of 3 weeks.

2.8.3 Phage titering

LB-tetracycline (10 ml) was inoculated with ER2738 from the streaked plate described previously and incubated at 37°C with shaking until OD_{600nm} reached approximately 0.5. The culture was then aliquoted into 200 µl fractions and infected with 10 ml of 10 to 10^2-fold serial dilutions of phage (prepared in 200 µl LB). Each aliquot was vortexed and incubated for 5 min at room temperature. The infected cells were added to 3 ml melted LB-Agar, vortexed and immediately poured onto freshly prepared LB-Agar containing IPTG/Xgal plates (0.05 µg/ml IPTG and 0.04 µg/ml Xgal [5-bromo-4-chloro-3-indolyl-β-D-galactosidase]). After setting, the plates were incubated at 37°C overnight with inversion. The following day, blue plaques were counted to determine the phage titre of each sample.

2.8.4 High-throughput sequencing of phage-displayed peptides

PCR was performed on the eluted or amplified phage samples using primers designed for Illumina HiSeq 2000 multiplexed paired-end read sequencing (Table 2-10).

Table 2-10: Phage display sequencing primers

Primer	***Sequence 5'-3'***	***Sample***
Forward 13	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACTCGCAATTCCTTTAGTGGTACC	Control FR1a
Forward 14	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACGCGCAATTCCTTTAGTGGTACC	Control FR1ua
Forward 15	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCCGCAATTCCTTTAGTGGTACC	Control FR2
Forward 16	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTAACGCAATTCCTTTAGTGGTACC	Control SR1a
Forward 17	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATCGCAATTCCTTTAGTGGTACC	Control SR1ua
Forward 18	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTAGCGCAATTCCTTTAGTGGTACC	Control SR2
Phd_Forward 25	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTCGCAATTCCTTTAGTGGTACC	Znf350 FR1a
Phd_Forward 26	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGCGCAATTCCTTTAGTGGTACC	Znf350 FR1ua
Phd_Forward 27	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCCCGCAATTCCTTTAGTGGTACC	Znf350 FR2
Phd_Forward 28	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCACGCAATTCCTTTAGTGGTACC	Znf350 SR1a
Phd_Forward 29	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTCGCAATTCCTTTAGTGGTACC	Znf350 SR1ua
Phd_Forward 30	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCGCGCAATTCCTTTAGTGGTACC	Znf350 SR2
Reverse	ACACGACGCTCTTCCGATCTCTGTATGGGATTTTGCTAAAC	

Key: Red = Illumina adaptor; Green = sequencing key for each sample; Black = phage DNA sequence

PCR reactions were set up on ice using Herculase II Fusion DNA polymerase (Agilent Technologies), as described below

1 µl Herculase II Fusion DNA polymerase
1X Herculase II Fusion DNA polymerase buffer
0.5M Betaine solution
13.2 nM Trehalose
500 µM dNTP
500 nM Forward primer
500 nM Reverse primer
1-10 µl eluted/amplified phage
Nuclease-free water up to 50 µl

Thermal cycling conditions:
95°C for 1 min
95°C for 15 sec
55°C for 20 sec
70°C for 1 min
Repeat steps 2-4 30 times
70°C for 3.5 min
Hold at 4°C

The entire PCR reaction was loaded on to a 2% (w/v) agarose gel and the amplicon of the correct size was extracted and purified using the QIAquick Gel Extraction Kit (Qiagen), according to the manufacturer's instructions. Purified DNA was eluted with 40 µl nuclease-free water and DNA concentration quantified using the Quant-iT PicoGreen dsDNA kit (Life Technologies), as per the manufacturer's instructions. Equal quantites of each amplicon sample were pooled and sent to Otogenetics, USA for sequencing. Sequencing data was obtained as fastq files and extracted in collaboration with Stuart Aitken then analysed using Microsoft Excel.

Chapter 3: Examining the interaction between IRF-1 and ZNF350

3.1 Introduction

IRF-1 has wide ranging roles in the immune response, cell proliferation, apoptosis, tumour progression plus invasion and metastasis [72; 73; 110; 111; 116; 154]. IRF-1 is traditionally known as a transcriptional activator where it stimulates the expression of genes which then induce various cellular processes. However, with the observation that IRF-1 represses *Cdk2* gene expression [99] there is an apparent lack of understanding how it can distinguish between activating or repressing gene transcription. It has been shown that proteins act in complexes at the gene promoter to exert transcriptional regulation so by identifying IRF-1-interacting proteins we may elucidate how IRF-1's transcriptional regulation activity is determined. In a previous study by the Ball group, KAP1 has been found to interact with the Mf2 domain of IRF-1. As previously mentioned, KAP1 is a transcriptional repressor that interacts with the KRAB domain of KRAB-containing zinc finger proteins such as ZNF350, and acts as a scaffold to recruit various proteins to the promoters of target genes resulting in gene repression [38-40]. Interestingly, other earlier studies by the Ball group have suggested that ZNF350 could also be an interacting partner of IRF-1. This hypothesis results from a yeast-2 hybrid screen using IRF-1 as the bait protein, which pulled out ZNF350 as a possible interacting protein. In addition, a peptide phage display screen identified a homologous ZNF350 peptide as a binding partner of the C-terminal region of IRF-1 (Figure 3-1). The majority of knowledge about ZNF350 originates from its interaction with BRCA1 [43; 45; 52; 54] and their activity as transcriptional repressors. This highlights a potential mechanism for IRF-

Peptide number	Homologous protein	Phage display peptide sequence	Domain	Blast peptide sequence
30	ZNF350	MVAPRTSL	256-325	VPS**VAP**Q**TSL**NI

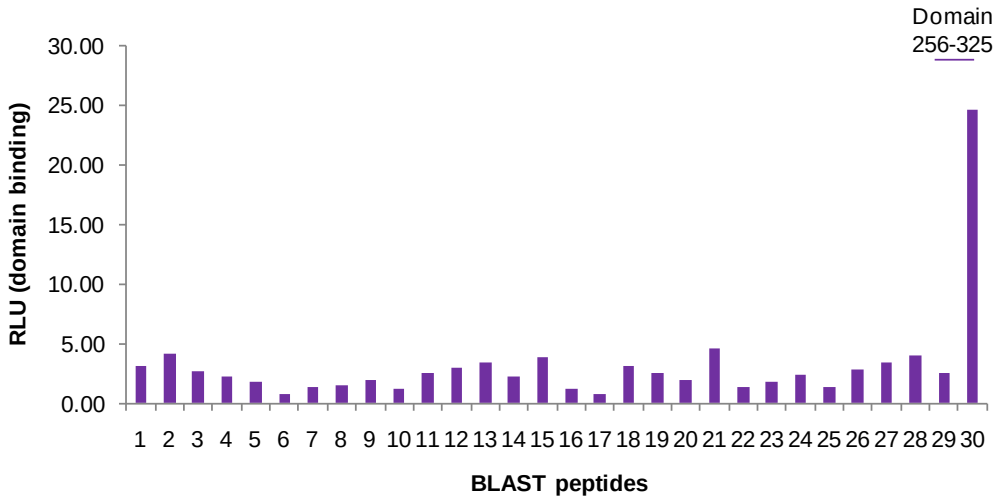


Figure 3-1: IRF-1 domain binding to peptides.

Performed by Angeli Möller [160]. Biotinylated 12-mer peptides were synthesised that encompassed the protein sequence homologous to the phage-display peptide that was shown to bind IRF-1. Immobilised peptides were incubated with each IRF-1 domain (1µg/ml). Following extensive washing, domain binding was detected with an anti-His antibody and enhanced chemiluminescence.

1-mediated transcriptional repression since its interaction with ZNF350 could facilitate its activity as a repressor as opposed to an activator. For this reason, I sought to determine if ZNF350 and IRF-1 interact both *in vivo* and *in vitro* and the circumstances under which the two proteins may interact

3.2 Results

3.2.1 Verification of the interaction between IRF-1 and ZNF350 in vivo

Although preliminary background data from previous Ball lab studies suggested that ZNF350 could be an IRF-1 partner protein, those observations had not been followed up and the interaction had not been validated. I therefore began by asking if IRF-1 and ZNF350 interact in a cellular environment. First, I set out to see if the endogenous proteins co-immunoprecipitated. Since A375 cells express wild-type IRF-1 and ZNF350 at readily detectable levels, I incubated A375 lysate with anti-ZNF350 and anti-IRF-1 antibodies which had been coupled to Protein G sepharose beads using the dimethyl pimelimidate (DMP). I then isolated and washed the beads before analysing the bound proteins by SDS-PAGE/immunoblots with anti-ZNF350 or anti-IRF-1 antibodies as stated (Figure 3-2A). Figure 3-2B shows that when the bound proteins were analysed using an anti-IRF-1 antibody to detect IRF-1 protein by immunoblot there was as expected IRF-1 in the α-IRF-1 pull-down whereas no IRF-1 protein was detected in a bead only control (beads-Ab). Strikingly, IRF-1 protein was also pulled-down by the anti-ZN350 antibody and indeed, with more efficiency than with the anti-IRF-1 antibody. The reason for this is unclear, however it may be that the C-terminal region of endogenous IRF-1 recognised by the anti-IRF-1 antibody could be inaccessible in some pools of IRF-1, for example due to its interactions with other proteins. This may be overcome by detecting the presence of IRF-1 when it is bound to another protein such as ZNF350. When the experiment was analysed using α-ZNF350 antibody to probe the immunoblots, I was able to show that a small but reproducible amount of ZNF350 was co-immunoprecipitated by the IRF-1 antibody, whereas no ZNF350 was seen in the bead control (Figure 3-2C). It should be noted that double bands for IRF-1 and ZNF350 are observed in both pull-down experiments (Figure 3-2B, lane 2 and 3-2C lanes 1 and 2, respectively). This could be due to different variations of the proteins being deteceted, for example truncated or post-translationally modified forms of the proteins. The data suggests that IRF-1 and ZNF350 form a complex in cell lysates

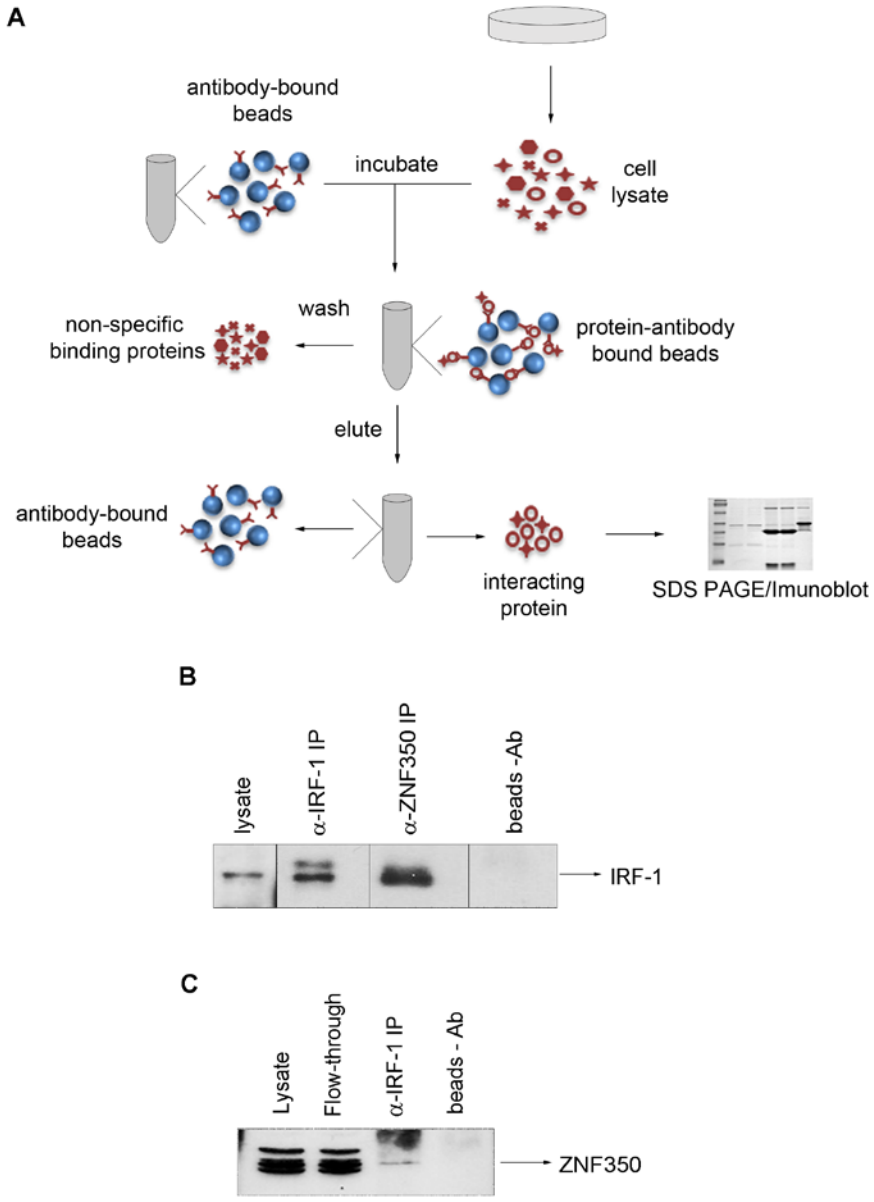


Figure 3-2: ZNF350 is found in a complex with IRF-1 in A375 cells

(A) Schematic depicting the method used to verify the interaction between IRF-1 and ZNF350 by immunoprecipitation from mammalian cell lysate using anti-IRF-1 and anti-ZNF350 antibodies. (B) A375 cell lysate (4 mg total protein/reaction) was incubated with protein-G beads alone (beads –Ab) or with protein-G beads that had been cross-linked to anti-IRF-1 Ab (α-IRF-1 IP) or anti-ZNF350 Ab (α-ZNF350 IP). Following extensive washing, bound proteins were analysed by 10% SDS-PAGE immunoblot using anti-IRF-1 mAb. Samples were from same experiment and run on a different gel. (C) Immunoprecipitation performed as in (B) using protein-G beads cross-linked to anti-IRF-1 Ab and analysed by 10% SDS-PAGE immunoblot using anti-ZNF350 polyclonal antibody.

and that although the complex appears to be captured more efficiently using α-ZNF350, it can also be detected in IRF-1 immuno-complexes.

3.2.2 Identification of the ZNF350 binding interface in IRF-1

Having demonstrated that IRF-1 and ZNF350 could be isolated in the stable complex, I next wanted to characterize the interaction in more biochemical detail. When the C-terminal enhancer domain of IRF-1 was used in a peptide phage display, one of the potential interacting peptides was found to be highly similar to a sequence within the ZNF350 protein (Figure 3-1). This suggested that ZNF350 might bind to the enhancer domain of IRF-1. In order to further elucidate the binding interface between the two proteins, I therefore used mutant IRF-1 protein constructs that were missing either the whole enhancer domain (Δenhancer) or the Mf1 sub-domain (ΔC25) (Figure 3-3A). I transfected One-strep tagged wild-type and mutant IRF-1 proteins into A375 cells (Figure 3-3B). As my preliminary studies, in agreement with previous Ball lab results, showed that IRF-1 complexes can be very transient and low affinity, I included the intracellular cross-linker DSP in this investigation to stabilize IRF-1:ZNF350 complexes. When the transfected A375 cells were lysed and One-Strep IRF-1 was captured using streptactin beads then analysed by SDS-PAGE/immunoblot (following reversal of the cross-linker), ZNF350 was shown to bind to wild-type IRF-1 (Figure 3-3B, lane 3) but not the Δenhancer mutant of IRF-1 (Figure 3-3B, lane 7). Interestingly, ZNF350 interacts with the Δ25 IRF-1 mutant with a reduced affinity compared to wild-type IRF-1 (Figure 3-3B, lanes 3 and 5), suggesting the last 25 amino acids are required for maximal ZNF350 binding. Thus, the Mf1 domain and the enhancer domain of IRF-1 may be required for direct binding to ZNF350 or alternatively, these domains may be required to maintain IRF-1 in a conformation to enable successful ZNF350 binding. It should be noted however, the IRF-1 is quite overloaded so it is difficult to deduce whether there is equal quantities of protein in each sample. To confirm this result with certainty, further investigations would be required.

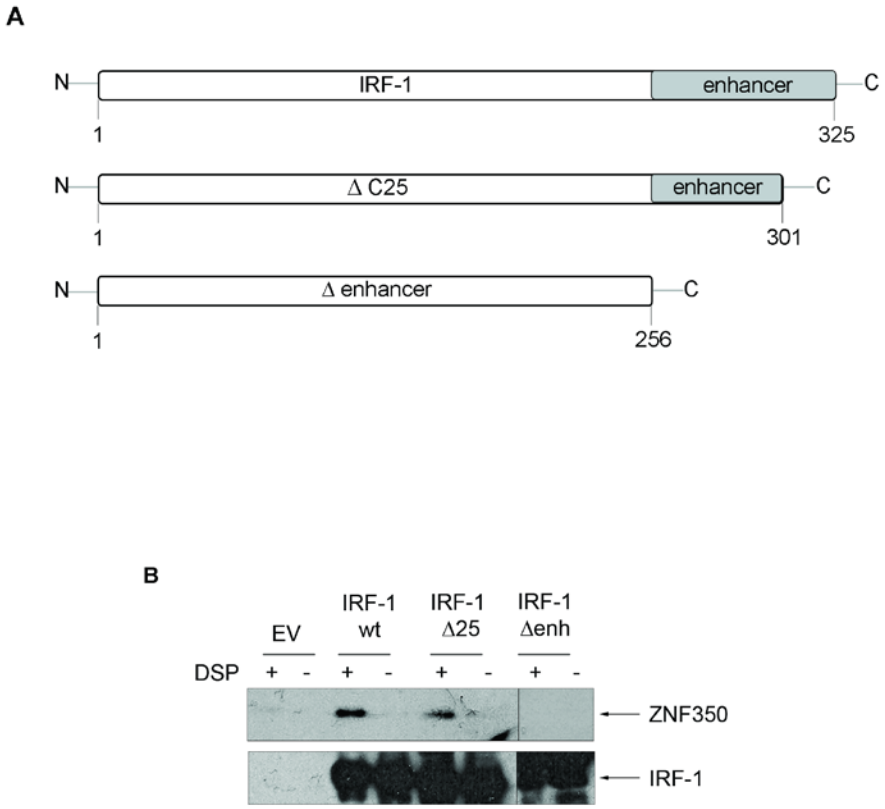


Figure 3-3: ZNF350 binds to the enhancer region of IRF-1

(A) Schematic depicting the One-Strep IRF-1 wild-type, ΔC25 and Δenhancer mutant constructs. (B) Immunoblot of ZNF350 immunoprecipitated with OneStrep IRF-1 and isolated using streptactin from A375 cells. Cells had been transiently transfected with OneStrep-IRF-1 wt, OneStrep-IRF-1 Δ25, OneStrep-IRF-1 Δenh or empty vector, with or without DSP crosslinker as indicated. The immunoblots were probed for ZNF350 and IRF-1.

3.2.3 Purification of recombinant His-ZNF350

In order to determine whether the interaction between IRF-1 and ZNF350 is direct or requires other cellular proteins, I used an *in vitro* approach to analyse the binding interface in more detail. First, I set out to purify ZNF350 protein from *E.coli*. There is no published data relating to the purification of full-length ZNF350 protein so I started out using two His-tag vectors, the standard T7 expression plasmid pET15b and the cold-shock expression vector pCOLD. The pCOLD expression system works by inducing protein expression at low temperatures to increase protein solubility and is known to aid in the production of difficult to express proteins that cannot be expressed in the T7 system. With no detectable ZNF350 expressed using the pET-15b vector, the pCOLD system provided a better starting point as some expression was detected, however the protein was highly degraded and the levels of expression were low. After initially expressing pCOLD ZNF350 and finding high levels of degradation, I decided to test expression and degradation at various time points. When expression trials were set up over a 4 h induction period, I found that 15 min was optimum (Figure 3-4A). Longer expression times result in less full-length His-ZNF350 and enhanced degradation products. I decided to use two types of *E.coli* cells to identify the best host for ZNF350 protein expression, BL21 DE3 and BL21 DE3 Star, the latter of which is designed for high level expression. To ensure the most efficient lysis of *E.coli* cells and subsequent purification of recombinant ZNF350, I employed a series of freeze-thaw cycles to disrupt the cell membrane. Briefly, after the *E.coli* cells had been induced to express the protein and collected by centrifuging, they were snap-frozen then resuspended in lysis buffer. Following incubation on ice for 30 min, the cells were once again snap-frozen then rapidly thawed in water bath at room temperature. This freeze-thaw cycle was repeated, after which the cells were briefly sonicated using a Soniprep 150 tip sonicator (Sanyo). Since ZNF350 could have been insoluble, I tested four different lysis buffers to ensure as much His-ZNF350 was extracted as possible (Figure 3-4B). Recipes of the lysis buffers used can be found in 2.5.1. Figure 3-4B shows that all four lysis buffers appear to extract comparable levels of protein after the first round of lysis in DE3* cells however, in DE3 cells lysis buffer 4 is only able to extract low

levels of protein. Interestingly, the re-extraction step results in a considerable increase in the levels of His-ZNF350 obtained when using lysis buffer 1, 2 or 4 while re-extraction with lysis buffer 3 appears to be unsuccessful in extracting any more soluble protein from the *E.coli* cells (Figure 3-4B). Lysis buffer 1 is able to extract the most protein in both the original lysis and the re-extraction so I decided to use this buffer for His-ZNF350 purification (Figure 3-4B, lanes 1, 5, 9). Interestingly, DE3* *E.coli* cells appear to express more His-ZNF350 than DE3 cells however, it is still apparent that most of this is insoluble and protein remains in the pellet (Figure 3-4B). In an attempt to increase levels of extraction in subsequent protein purifications, the *E.coli* cells were mechanically lysed using a bioruptor, allowing for more efficient break down of the cell wall and more soluble protein obtained in the cell lysate.

Initially, His-ZNF350 would not couple to the Ni^{2+}-NTA agarose beads used for purification of His-tagged proteins and increasing the ratio of nickel beads to lysate did not alleviate this problem so I hypothesised that the His-tag may be 'cryptic', meaning it may possibly be masked by the folded protein. To attempt to counteract this problem I tested different concentrations of urea in both the lysis buffer and wash buffers since urea helps solubilise the protein and unfold it to expose the His-tag (Figure 3-5). While ZNF350 may initially need to be denatured to allow for binding to Ni^{2+}-NTA agarose beads, ideally the protein should be in its folded state for use in experiments for the results to best reflect the behaviour of ZNF350 in its native conformation. In addition, urea can interfere with subsequent assays for analysis of ZNF350 such as ELISAs thus His-ZNF350-bound Ni^{2+}-NTA beads were washed with buffers in a step-wise reduction of urea to gently re-nature the protein and get rid of the urea in the buffer (8M, 6M, 4M, 2M and 0M; Figure 3-5A-E). To reduce non-specific binding of other His- proteins to the nickel beads, the lysis buffer and wash buffers contain a low concentration of imidazole, which competes with His- for nickel binding. Despite working at $4^{o}C$ and using protease inhibitors in all buffers, His-ZNF350 protein appears degraded under most conditions, however it can be noted that these degradation products are gradually lost after multiple rounds

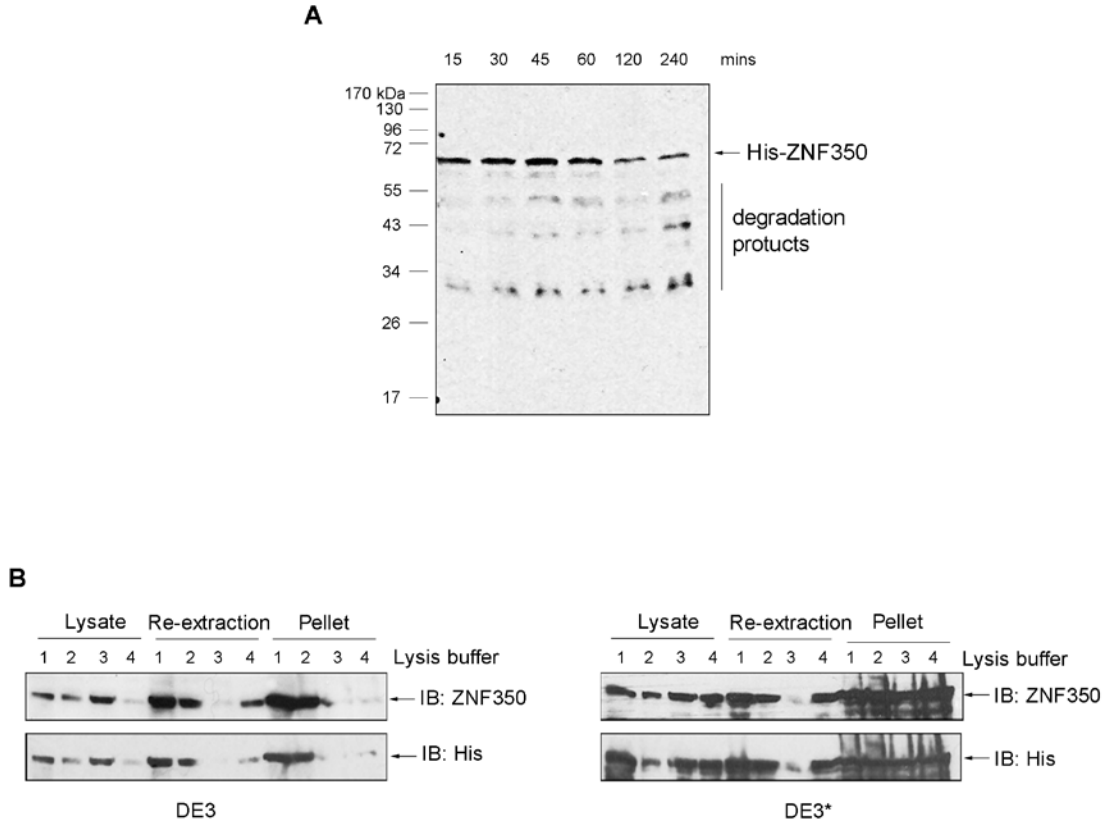


Figure 3-4: Expression of ZNF350 in E.coli

(A) Immunoblot of His-ZNF350 following different lengths of induction by IPTG (0.5 MM) in BL21 DE3 cells and probed with anti-ZNF350 antibody. (B) BL21 DE3 and (C) BL21 DE3* cells were used to compare levels of ZNF350 using different lysis buffers (1-4, recipe detailed in Materials and Methods). ZNF350 was detected in lysates, re-extraction and pellets by SDS-PAGE/immunoblot and developed using anti-ZNF350 and anti-His antibodies.

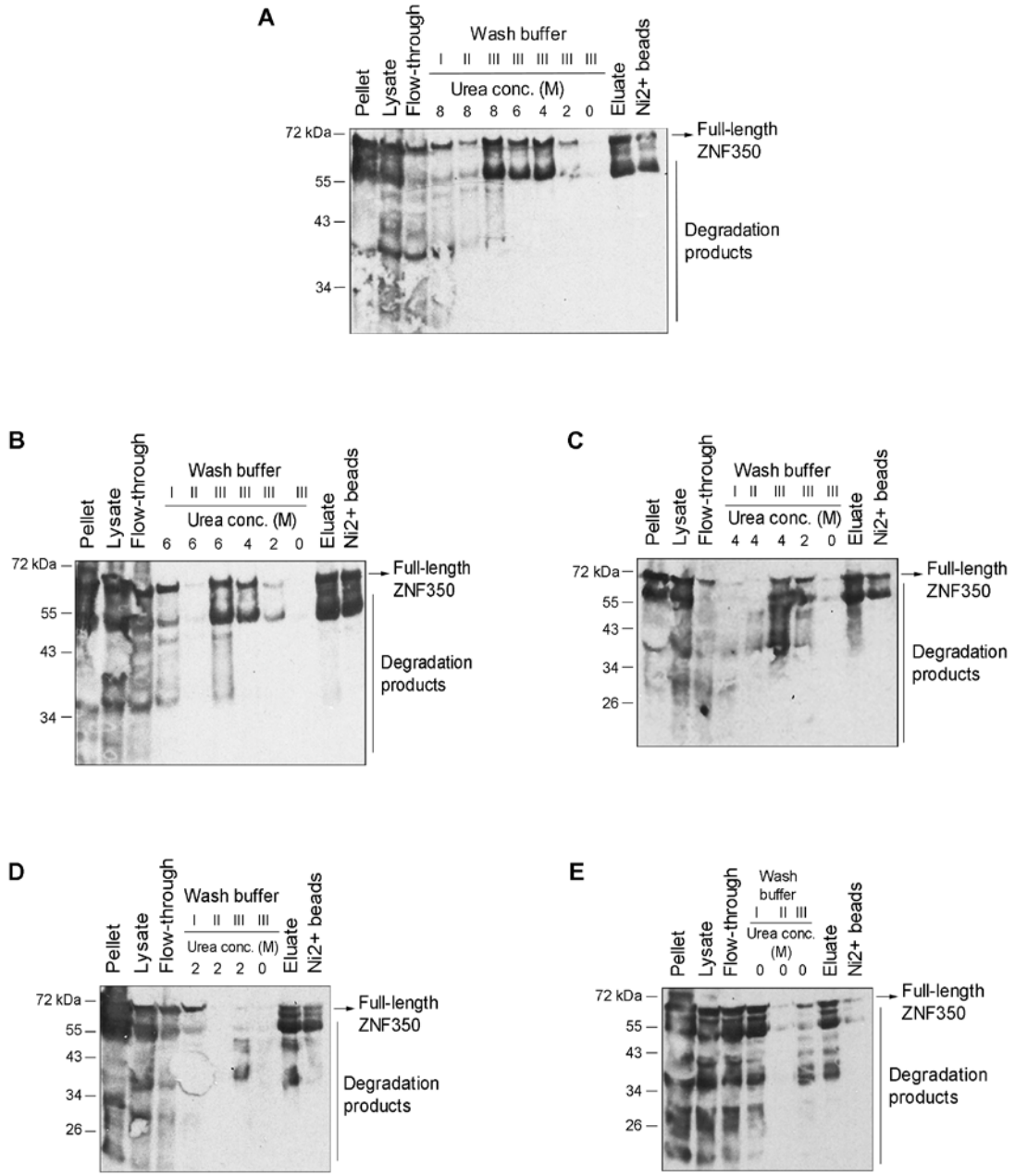


Figure 3-5: His-ZNF350 purification with urea buffers

His-ZNF350 was lysed in buffer containing (A) 8M urea (B) 6M urea (C) 4M urea (D) 2M urea or (E) 0M urea and washed in buffers with a step-wise reduction in urea concentration (8M to 0M, as indicated). The levels of protein were detected using SDS-PAGE/immunoblot and probed using anti-ZNF350 antibody.

of column washing and as the concentration of urea in the buffers is reduced, suggesting that the majority are N-terminal deletions missing the His-tag (Figure 3-5).

After eluting His-ZNF350 from Ni^{2+}-NTA agarose beads using a high concentration imidazole buffer, there appeared to be a lot of protein remaining bound to the beads. Increasing the imidazole concentration did not increase elution efficiency so I increased the incubation time of the elution buffer with the His-ZNF350 bound nickel beads to 24 hours. On occasion, this resulted in a small amount of protein being eluted however, most of the protein appeared to elute from the beads within 2 hours (Figure 3-6A). His-ZNF350 is eluted from Ni^{2+}-NTA agarose beads using a buffer containing a high concentration of imidazole which can interfere with protein quantification assays so I sought to exchange the protein buffer to omit imidazole. After dialysis of the protein resulted in protein precipitation, I used Amicon ultra devices to concentrate and buffer exchange His-ZNF350. Figure 3-6B demonstrates that when buffer exchanging using the Amicon filters, a lot of protein is lost, presumably through precipitation and concentrating the volume of eluate does not result in an increased concentration of His-ZNF350. It should be noted however that the His-ZNF350 eluate can be concentrated using the Amicon columns when in imidazole buffer but as soon as attempts to exchange the buffer are made, the protein levels decrease considerably. Eventually, I buffer exchanged His-ZNF350 very slowly using a HiTrap desalting column on an AKTA chromatography system to obtain protein in imidazole-free buffer. To determine the His-ZNF350 concentration, I precipitated the protein using acetone and analysed the concentration by comparison to BSA standards (Figure 3-6C). The final ZNF350 preparation was analyzed by silver stained SDS-PAGE gel complemented by immunoblot. Although the concentration and buffer exchange steps lead to further degradation of ZNF350, I was satisfied that the protein would be of sufficient quality to determine if it bound directly to IRF-1 *in vitro*.

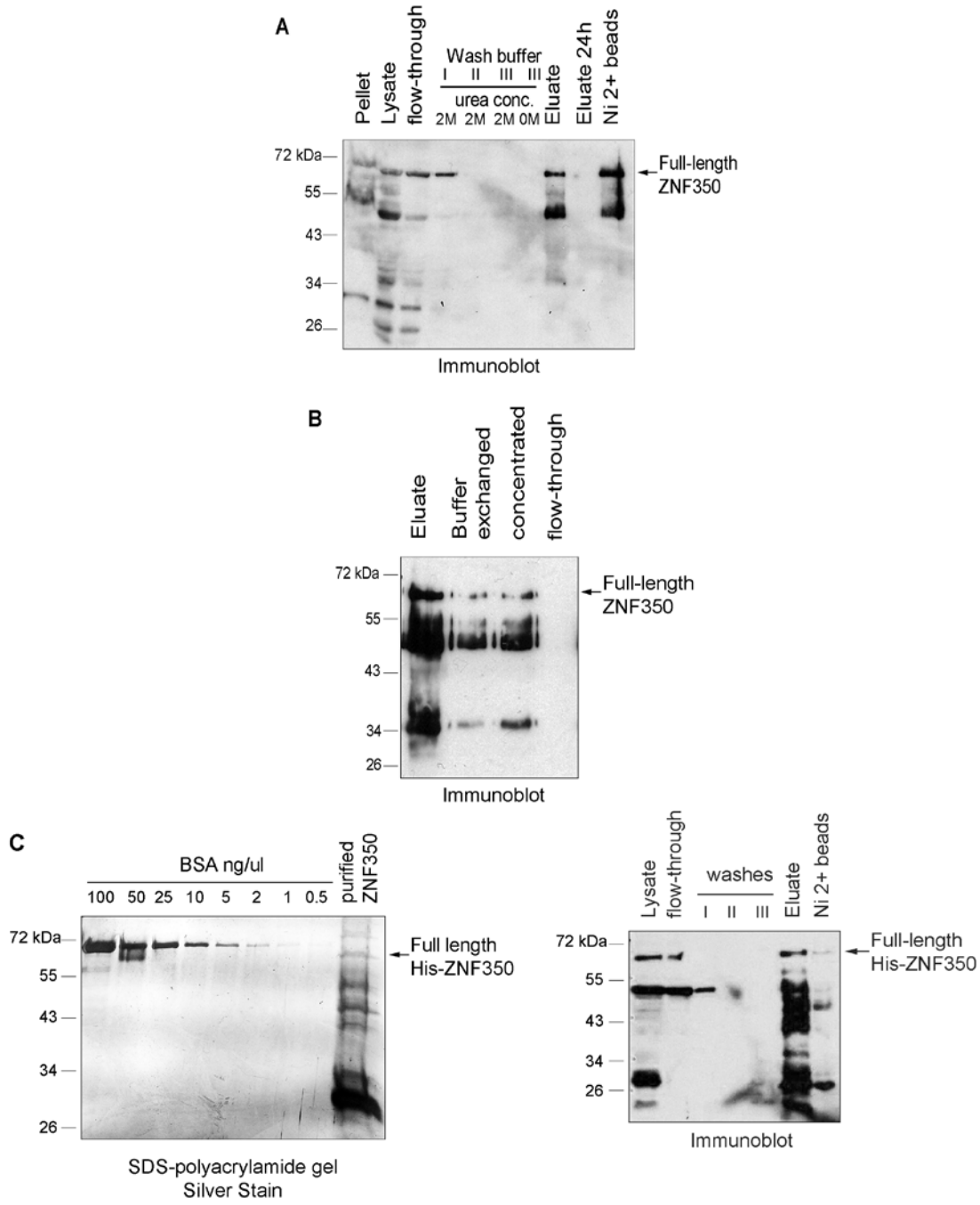


Figure 3-6: His-ZNF350 elution and buffer exchange

(A) Immunoblot of His-ZNF350 purified from BL21 DE3* cells and incubated with elution buffer for 2 h or 24 h and protein analysed by SDS-PAGE/immunoblot and detected using anti-ZNF350 antibody. (B) His-ZNF350 levels detected following buffer exhange and concentration using Amicon fliters and analysed as above. (C) Full-length His-ZNF350 detected by silver stained SDS-PAGE gel and compared to BSA standards (left panel). Immunoblot of optimised purification of His-ZNF350 detected as described above.

3.2.4 Confirmation of ZNF350 binding to IRF-1 in vitro

Having successfully purified His-ZNF350 from *E.coli* cells, I sought to normalise the protein concentration. By adding a titration of His-ZNF350 to a microtitre plate, I was able to detect the presence of ZNF350 with an α-ZNF350 antibody in a concentration-dependent manner (Figure 3-7A). This ultimately shows that the His-tagged protein purified is in fact ZNF350. To establish whether ZNF350 and IRF-1 interact *in vitro*, I immobilised recombinant GST-IRF-1 on a microtitre plate and incubated it with a titration of His-ZNF350 in the mobile phase. I found that ZNF350 could interact with full-length IRF-1 in a dose-dependent manner (Figure 3-7B), which supports the previous studies showing an interaction between the two proteins *in vivo* and extends it by showing that this interaction is likely to be direct.

The One-Strep IRF-1 pull-down experiments shown in Figure 3-3 suggested that ZNF350 might bind directly to the extreme C-terminus of IRF-1 (Mf1 domain) or that this domain might be indirectly involved in ZNF350 binding. To clarify which of these mechanisms are responsible for the interaction with ZNF350, I used a peptide-binding assay. In this assay a biotin labelled peptide based on the Mf1 domain (C-terminal 25 aa) was displayed on a microtitre plate by streptavidin capture alongside a control peptide from a sequence in the p21 protein (Figure 3-7C). When the peptides were incubated with a titration of ZNF350 and binding detected using an anti-His antibody, the protein bound to the Mf1 domain peptide whilst no significant binding to the p21 peptide was detected. Thus the C-terminus of IRF-1 contains a ZNF350 docking site.

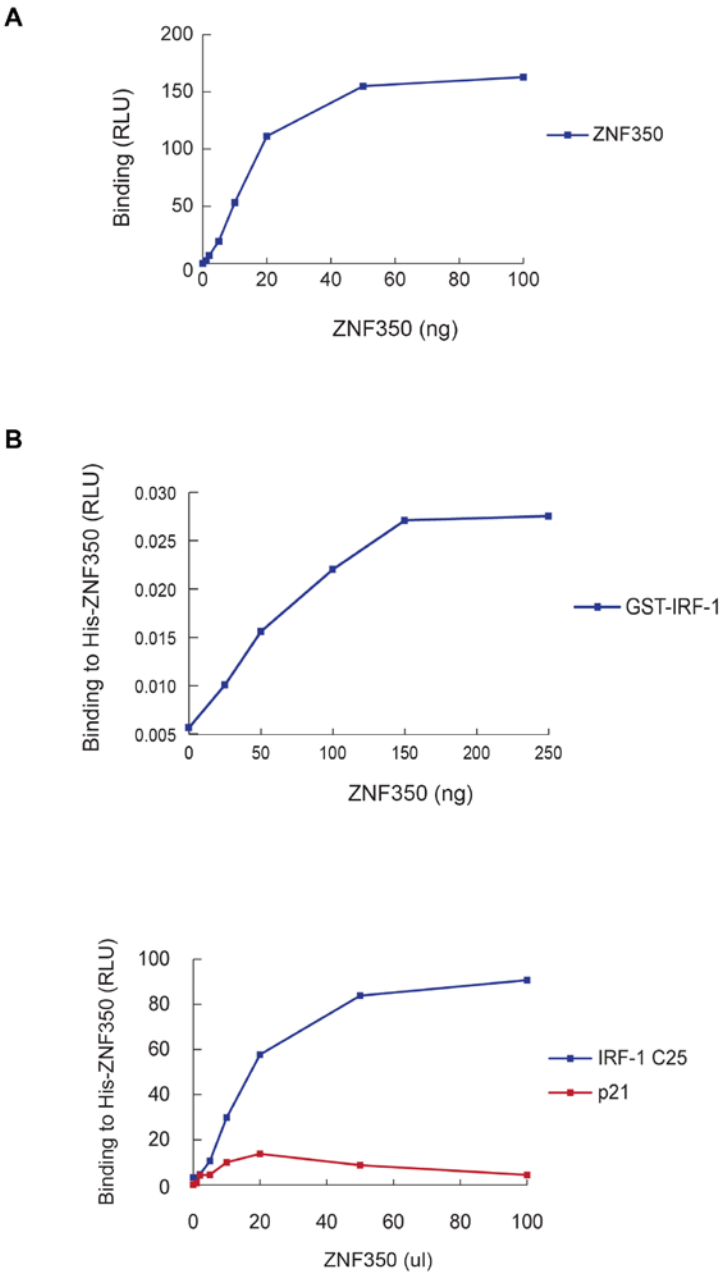


Figure 3-7: His-ZNF350 interacts with IRF-1 in vitro

(A) A titration of recombinant His-ZNF350 (0-100 ng) was coated on a microtitre plate and detected using anti-His mAb and enhanced chemiluminescence. Amount of protein, expressed as relative light units (RLU) is shown. (B) Recombinant GST-IRF-1 (100 ng) was coated on a microtitre plate, incubated with a titration (0-250 ng) recombinant His-ZNF350 and binding detected using anti-IRF-1 mAB. (C) Saturating levels of biotin-tagged IRF-1 C25 or p21 peptide were added to a microtitre plate coated with streptavidin and incubated with a titration of recombinant His-ZNF350 (0-100 µl). Binding was detected as in (B).

3.2.5 Fine-mapping of the ZNF350 binding interface in IRF-1

Having identified one ZNF350 binding interface in the C-terminus of IRF-1, I next wished to establish if other interfaces were also involved in binding. In order to further characterise the interaction between ZNF350 and IRF-1, I used a method developed within the Ball group which employs a panel of IRF-1 overlapping peptides coupled to streptavidin-agarose columns and is used to capture endogenous cellular ZNF350 rather than recombinant protein to establish specific regions of IRF-1 responsible for binding. The IRF-1 peptides are 20 aa long with an N-terminal biotin tag, a Ser-Gly-Ser-Gly spacer and a shared 5 aa overlap with the successive peptide (Figure 3-8A). Briefly, each biotinylated peptide was immobilised on streptavidin-agarose to which A375 cell lysate was added. The cell lysate was pr-cleared with Sepharose-4B beads prior to addition to the columns to get rid of 'sticky' proteins that may bind to the beads matrix as opposed to specifically the IRF-1 peptides. After extensive washing, the bound proteins were eluted by adding beads to SDS-PAGE sample buffer and analysed using SDS-PAGE/immunoblot developed using an anti-ZNF350 antibody to identify binding of ZNF350 to specific IRF-1 regions (Figure 3-8B).

When using this technique, I found that ZNF350 binds stably to several IRF-1 peptides, namely peptides 7, 8 and 21 (Figure 3-9). ZNF350 also appears to bind with low affinity to peptide 9 which, along with peptides 7 and 8 (aa 91-140), span the Multi-functional 2 (Mf2) domain (aa 106-140) and encompass the nuclear localisation signal (aa 117-141). Characteristically of NLS sequences, this region of IRF-1 is rich in basic residues and the large number of charged residues suggests this region is likely to be exposed. Early studies using a polyclonal antibody raised to full-length IRF-1 in fact confirmed the presence of this polar region [82]. This Mf2 region of IRF-1 has also been shown to be highly disordered, which appears to aid its function as a multi-protein binding interface [87]. ZNF350 binds with the highest affinity to peptide 21 (aa 301-320), which spans most of the Multi-functional 1 (Mf1) domain (301-325), in agreement with the results obtained with recombinant ZNF350 (Figure 3-7). The data also suggests that the complete loss of ZNF350 binding seen

when the enhancer domain of IRF-1 is deleted (Figure 3-2) likely represents an effect on IRF-1 core domain conformation rather than a second enhancer domain binding interface. The Mf1 domain is required for maximal IRF-1-mediated growth suppression and plays a key role in determining the rate of IRF-1 degradation [98; 101]. Since results from previous studies showed that ZNF350 binding to IRF-1 is completely lost when the enhancer domain is deleted, this likely indicates that the C-terminal region is critical for the association of the two proteins.

In addition to the results obtained using peptide affinity columns, when IRF-1 peptides were immobilised on a microtitre plate, recombinant His-ZNF350 was found to bind to peptides within the Mf1 and Mf2 domains (data shown in Appendix I). However, the peptides His-ZNF350 bound to were peptide 9 (Mf2) and peptide 22 (Mf1) and although they are not the exact peptides bound by the endogenous protein, this may be explained by the intrinsically disordered nature of these two domains. Peptides 8 and 9 overlap by 5 residues while peptides 21 and 22 overlap by 15 residues so it may be that endogenous ZNF350 and recombinant ZNF350 recognise similar residues but in different peptides, possibly due to their different conformations.

As the Mf1 domain is known to contain a number of critical regulatory motifs [161], I next I set out to fine map the interaction between ZNF350 and the Mf1 domain peptide of IRF-1 to identify critical contact residues. A library of IRF-1 peptides based on the Mf1 domain in which each amino acid was sequentially replaced with an alanine residue were used to generate a series of affinity columns. Similar to the previous section, A375 cell lysate was loaded onto the columns and following extensive washing, bound proteins were eluted and analysed by immunoblot to identify peptides which showed reduced binding to cellular ZNF350. Using this approach, I found that ZNF350 binding to the Mf1 domain was lost when several residues are mutated. These include residues within the co-regulator signature LXXLL motif (aa 306-310) and a number of threonine and proline residues encircling the motif (Figure 3-10B). This indicates that the LXXLL motif within the IRF-1 C-terminal domain is very important for the interaction with ZNF350.

The LXXLL motif was originally detected in cofactor proteins that interact with hormone-activated nuclear receptors. However, these motifs have also been observed in other proteins such as the transcription factor STAT6 [162] and the transcriptional co-activators calcium response element-binding protein (CREB)-binding protein (CBP) and p300 [163]. In IRF-1, this LXXLL motif is required for binding of the nuclear chaperone Hsp70 family members, which cooperate with Hsp90 to regulate IRF-1 turnover and activity [161]. Furthermore, the LXXLL motif has been shown to be critical for IRF-1-dependent repression of *Cdk2* [98]. Interestingly, when some residues are mutated to alanine, an increased affinity for ZNF350 binding can be observed, namely Ser^{317} and Pro^{322} (Figure 3-10B) suggesting that ZNF350 binding has evolved to be low affinity and transient as might be expected for a regulatory interaction.

In an attempt to fine-map the interaction between ZNF350 and the Mf2 domain of IRF-1, I carried out a similar experiment using a library of IRF-1 peptides based on peptides 7 and 8, where each amino acid was sequentially mutated to an alanine. However, the results obtained were quite variable and no definitive conclusions could be drawn from the data. This may have been due to the intrinsically disordered nature of the Mf2 domain and the transient and varied interactions it participates in.

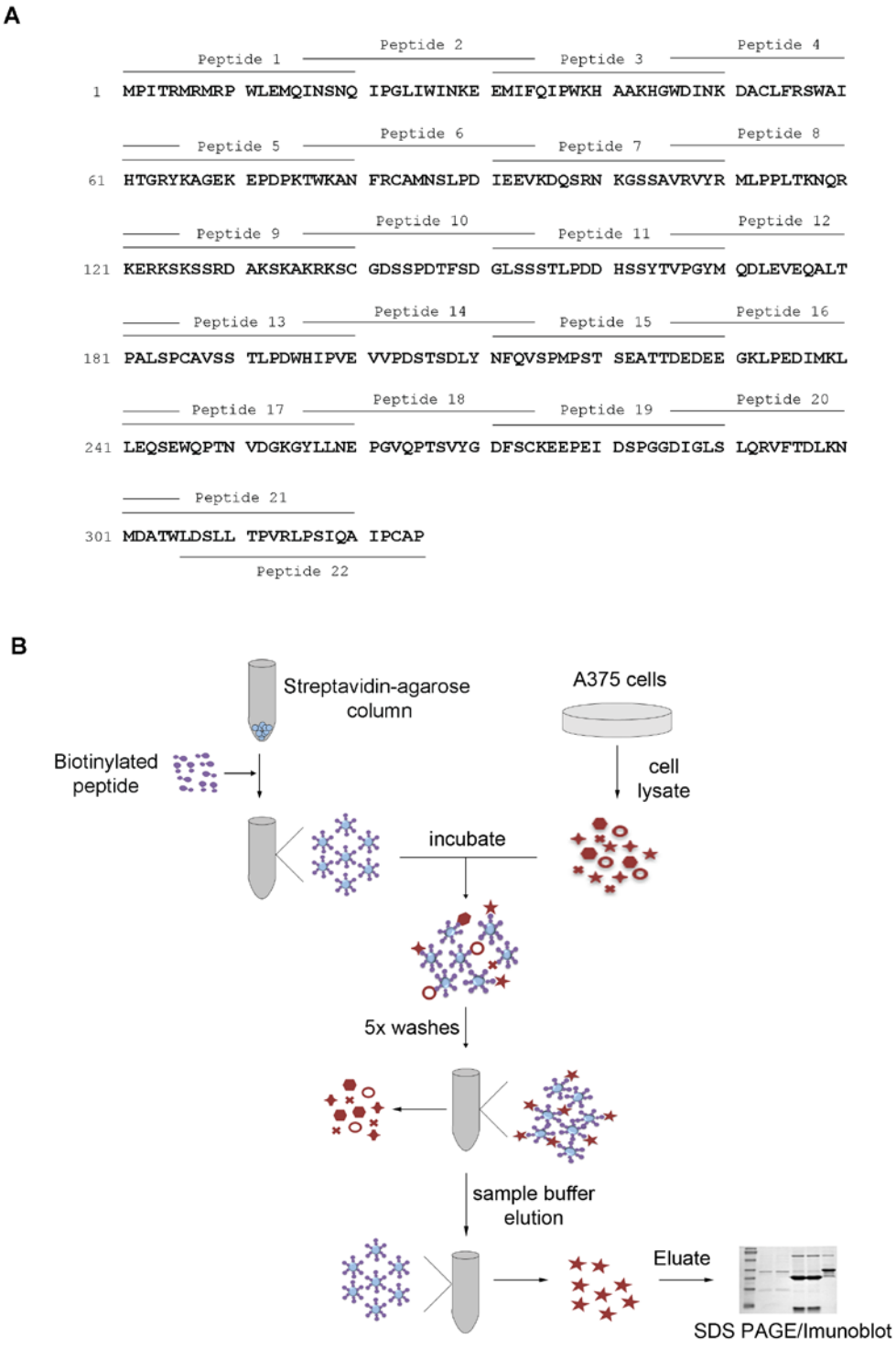


Figure 3-8: Method used to map the binding of ZNF350 to IRF-1

(A) 20 aa peptides with an N-terminal biotin tag and a 5 aa overlap spanning the entire IRF-1 protein sequence. (B) Schematic showing the method used to identify the region of IRF-1 to which endogenous ZNF350 interacts using overlapping IRF-1 peptides.

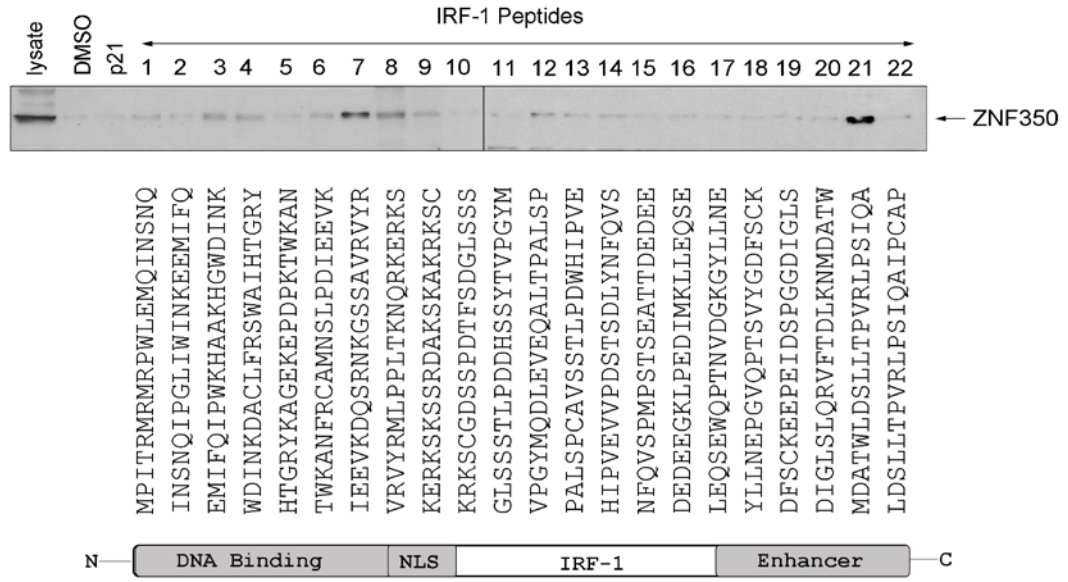


Figure 3-9: ZNF350 has multiple binding sites on IRF-1

(A) Eluates (20 µl) from the peptide affinity chromatography depicted in Figure 3-8B were run out on 10% SDS-PAGE gel and ZNF350 binding was detected by immunoblot using anti-ZNF350 antibody.

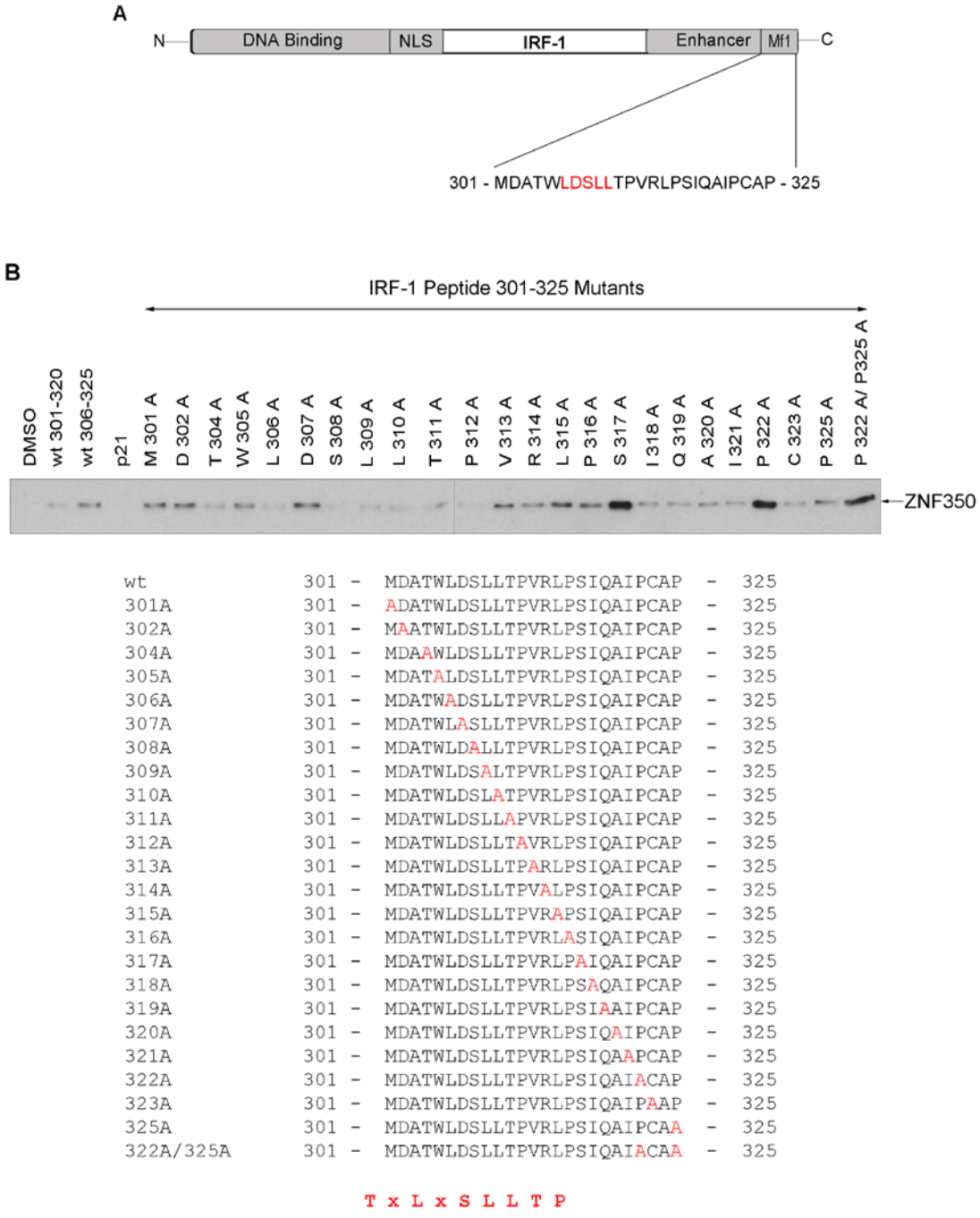


Figure 3-10: ZNF350 binds to the LXXLL motif in IRF-1

(A) Scematic depicting IRF-1 domains, specifically the Mf1 domain containing the co-regulator motif LXXLL. (B) A panel of peptides containing alanine substitutions so that each amino acid was sequentially mutated (lower panel) was used to generate a series of affinity columns. Following the isolation of peptide binding proteins from A375 cell lysate, the eluate was analysed by SDS-PAGE/immunoblot developed using anti-ZNF350 antibody.

3.2.6 Determining the consequence of cellular signals on ZNF350 and IRF-1 protein expression and interaction

The expression level of IRF-1 increases in response to viral infection, as well as in response to DNA-damaging agents [164]. IRF-1 can regulate specific target genes, thus inducing cell cycle arrest and/or apoptosis in response to genotoxic stress [114; 120]. Since IRF-1 and ZNF350 interact with one another, I was interested to see what effect IRF-1-inducing signals had on the expression of ZNF350 and consequently its association with IRF-1. I treated A375 cells with increasing concentrations of IFNγ or etoposide and lysed the cells using RIPA buffer, which enables the extraction of nuclear proteins such as ZNF350 and IRF-1 (Figure 3-11). Figure 3-11A illustrates that when cells are treated with increasing concentrations of IFNγ, the expression of IRF-1 protein is substantially increased even at low concentrations of interferon, whereas ZNF350 protein levels are only slightly amplified compared to untreated cells. When cells are treated with etoposide, there is a substantial dose-dependent increase in IRF-1 protein levels (Figure 3-11B). On the other hand, there is only a slight increase in levels of ZNF350 in etoposide-treated cells compared to untreated cells with higher concentrations of etoposide having no added effect on protein expression (Figure 3-11B). This marginal induction of ZNF350 expression in response to IRF-1-inducing agents is not directly correlated to the significant amplification of IRF-1 expression, indicating that the expression of the two proteins is independent of one another.

In order to determine if the changes in IRF-1 protein levels have an effect on its interaction with ZNF350, I treated A375 cells with IFNγ (100 U/ml) or etoposide (20 mM) (Figure 3-12A). As expected, IRF-1 levels are increased following treatment with both agents (Figure 3-12A, lower panel lanes 2 and 4, respectively) compared to their controls (Figure 3-12A, lower panel lanes 3, 5, respectively), whereas ZNF350 protein levels do not appear to be significantly affected by any of these signals (Figure 3-12A, upper panel). I analysed the effect these treatments had on the interaction by immunoprecipitating IRF-1 from treated mammalian cell lysates using

anti-IRF-1 antibody coupled to Protein G sepharose beads and detected the presence of bound ZNF350 (Figure 3-12B). Interestingly, the interaction with ZNF350 does not increase in response to higher levels of IRF-1 brought about by IFNγ treatment (Figure 3-12B, lanes 3 and 4) which indicates that the association of the two proteins is not induced by IFNγ. When cells are treated with etoposide, there appears to be a slight decrease in the interaction with ZNF350 despite the amplification of IRF-1 levels (Figure 3-12B, lanes 5 and 6). It proved quite difficult to identify ZNF350 following immunoprecipitation since reversing the crosslinks used to initially hold the two proteins together often proved unattainable, despite trying various different methods. When the eluted proteins were analysed using western blotting, the ZNF350 protein is roughly the same size as the IgG heavy chain from the antibody used in the pulldown. Despite using low percentage SDS-PAGE gels to separate the proteins out as much as possible, it was still difficult to distinguish ZNF350 from the antibody heavy chain. However, from the results it can be seen that despite a substantial increase in IRF-1 levels in response to the two agents, this does not result in an enhanced interaction with ZNF350 (Figure 3-12B).

Yun *et al*.[68] reported that ZNF350 is rapidly degraded upon treatment with DNA damage-inducing UV, so I sought to evaluate if this possible depletion of ZNF350 affected the association with IRF-1. A375 cells were exposed to increasing levels of UV, lysed 2 h after treatment and the lysates subsequently incubated with anti-IRF-1 antibody coupled to protein G sepharose beads to extract IRF-1 along with any bound proteins. ZNF350 protein levels appear to slightly decrease in response to higher levels of UV irradiation, specifically 50 j/m^2 (Figure 3-13A, upper panel). This may support the findings of Yun et al. who detected a loss of ZNF350 in response to UV. However, the authors demonstrated a more striking loss of protein but this may be explained by their use of ectopically expressed GFP-ZNF350 to detect changes in levels of protein, whereas the results displayed here are from the effects on endogenous ZNF350. IRF-1 protein expression on the other hand appears to be initially induced at low levels of UV but this then decreases slightly in response to higher levels of UV exposure (Figure 3-13A, lower panel). As a result, when ZNF350 is co-immunoprecipitated with IRF-1, the levels of association are

significantly decreased, probably due to the lower levels of protein (Figure 3-13B). Again, this finding demonstrates that the expression of the two proteins is independent of one another and their association is perturbed in response to a decrease in IRF-1 and/or ZNF350 protein levels.

Since ZNF350 does not appear to be induced in response to interferons or DNA damage in the same way IRF-1 is, I sought to explore if other effectors of IRF-1 activity altered ZNF350 expression. The transcription inhibitor 6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) has been shown to prevent IRF-1-induced gene expression and also significantly inhibits basal transcription [165]. DRB is a potent inhibitor of CDK7 and CDK9, kinases responsible for the phosphorylation of the C-terminal domain (CTD) of the RNA polymerase II large subunit [166; 167]. It inhibits more than 50% of mRNA synthesis above 40 μM [168; 169] and has been shown to inhibit both RNA Pol II CTD phosphorylation *in vitro* [166] and transcription *in vivo* [170]. Inhibition of RNA Pol II-dependent transcription, including that due to DRB, has been shown to trigger cell death, [168; 169; 171] however the exact mechanisms for this remain unclear. I treated A375 cells with increasing concentrations of DRB and found that ZNF350, IRF-1 and KAP1 levels all decrease in a dose-dependent manner, with a complete loss of IRF-1 after treatment with 50 μM DRB (Figure 3-14A). In contrast, p53 is induced in response to DRB (Figure 3-14A), an observation which has been reported previously [172; 173]. It has been shown that the p53 accumulated in response to DRB is not phosphorylated at Serine 15, a feature increased after cell damage. This suggests that the increase in p53 levels observed after DRB treatment is not due to the activation of the DNA damage response pathway by DRB [174]. Interestingly, DRB has been shown to severely reduce the levels of the p53 regulator protein MDM2, and in cells where MDM2 is completely lost after DRB treatment, p53 protein levels are significantly amplified [175]. It has also been found that E2F-1 is stabilized when MDM2 disappears following DRB treatment, in a manner similar to that observed with p53 [175]. This together with evidence from Martin et al. [176] for a physical interaction between MDM2 and E2F-1, suggest that both p53 and E2F-1 are targeted by MDM2 for degradation . E2F-1, in complex with Rb, CtIP and CtBP, has been

shown to directly bind to an E2F recognition sequence on the ZNF350 promoter and repress transcription [71] so it could be hypothesised that an accumulation of E2F-1 in response to DRB may promote a decrease in ZNF350 gene expression, although further studies would be required to verify this. DRB activity however is reversible and does not inhibit transcription elongation, which therefore allows transcription to continue until the RNA polymerases have completed their cycle on a particular gene. To understand the effect DRB has on ZNF350 expression in greater detail, I treated A375 cells with DRB and lysed the cells at various time points. Figure 3-14B shows that ZNF350 expression actually increases, the longer the cells have been exposed to DRB. Despite loading the same volume of cell lysate for each sample, it could be that the overall protein levels are not equal. Therefore a loading control, such at detecting GAPDH levels, would be required to identify if the increase in ZNF350 levels is true or not. Since ZNF350 seemed to diminish at high levels of DRB treatment, it would be interesting to carry out further studies to understand how ZNF350 may accumulate after long-term DRB treatment.

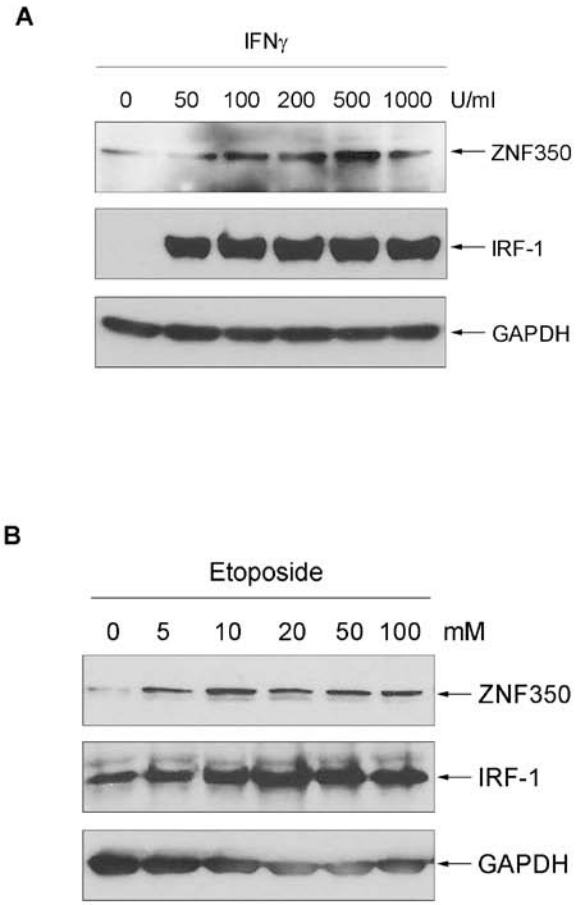


Figure 3-11: Effect of IRF-1-inducing agents on ZNF350 levels

A375 cells were treated with a titration of (A) IFNγ (0-1000 U/ml) or (B) etoposide (1-100 mM) and analysed by SDS-PAGE/immunoblot. Protein levels were detected using anti-ZNF350, anti-IRF-1 or anti-GAPDH antibodies, as indicated.

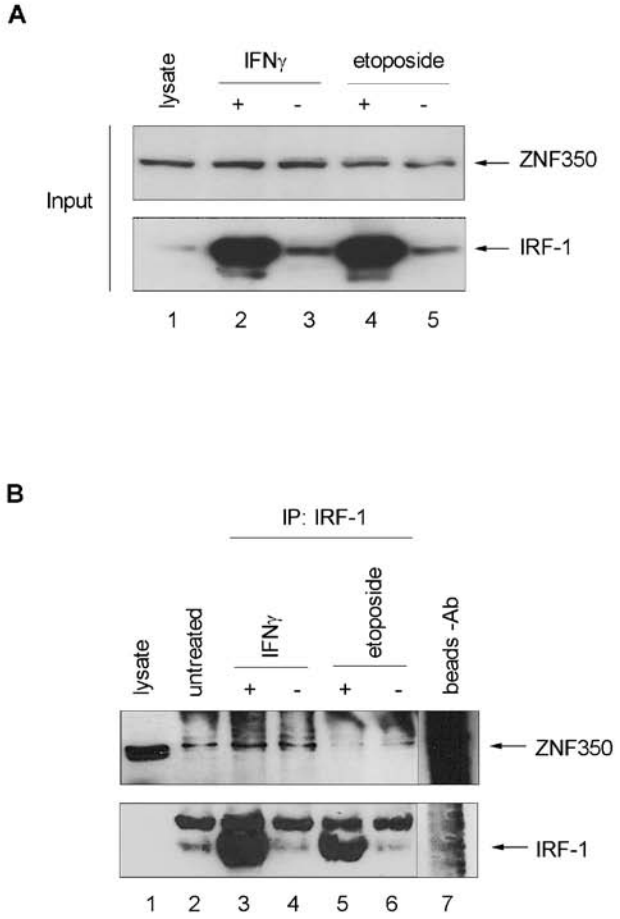


Figure 3-12: ZNF350 protein levels and the interaction with IRF-1 are not affected by cell signals

(A) A375 cells were treated with IFNγ (100 U/ml) or etoposide (20 mM) and analysed by SDS-PAGE/immunoblot. Levels of ZNF350 and IRF-1 were detected using anti-ZNF350 and anti-IRF-1 antibodies, respectively. (B) A375 lysates from (A) were incubated with protein-G beads alone (beads –mAb) or protein-G beads cross-linked to anti-IRF-1 mAb and binding detected by SDS-PAGE/immunoblot using anti-ZNF350 or anti-IRF-1 antibodies as indicated.

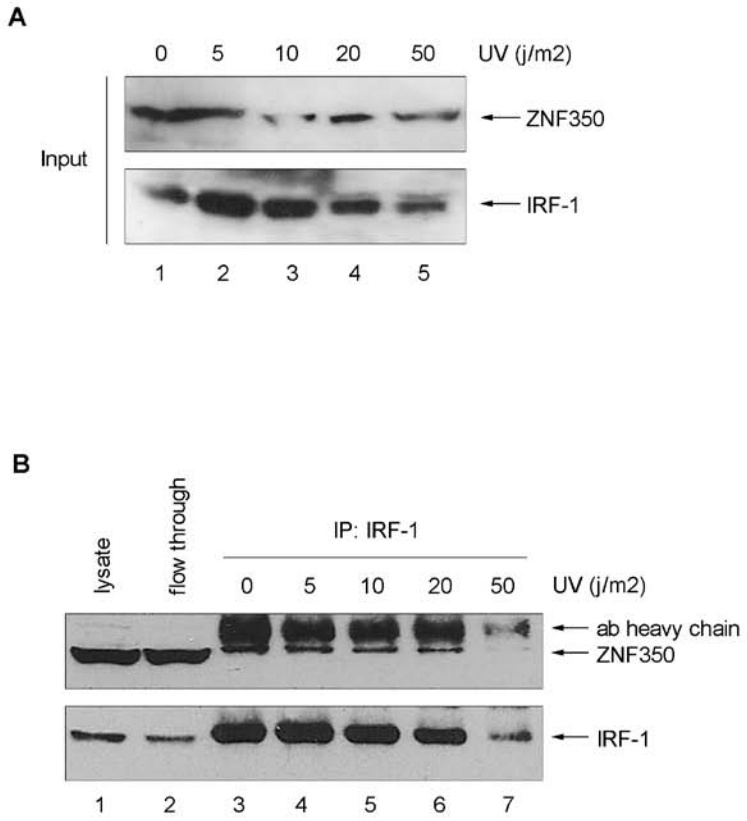


Figure 3-13: The interaction between ZNF350 and IRF-1 is reduced following UV treatment

(A) A375 cells were treated with a titration of UV (0-50 j/m2) and analysed by SDS-PAGE/immunoblot. Levels of ZNF350 and IRF-1 were detected using anti-ZNF350 and anti-IRF-1 antibodies, respectively. (B) Lysates from (A) were incubated with protein-G beads crosslinked to anti-IRF-1 mAb and binding detected by SDS-PAGE/immunoblot using anti-ZNF350 or anti-IRF-1 antibodies as indicated.

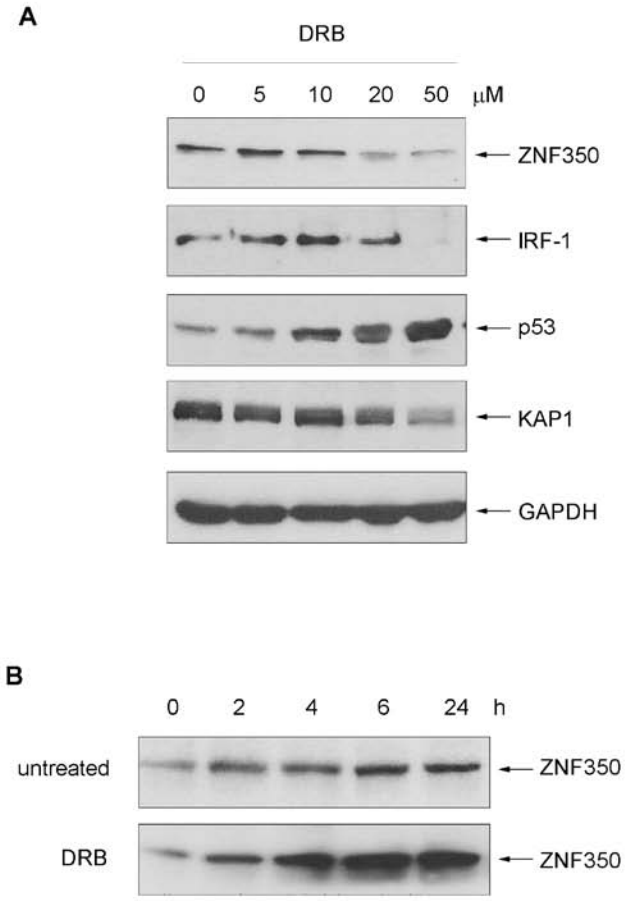


Figure 3-14: ZNF350 and IRF-1 levels are reduced following treatment with DRB

(A) A375 cells were treated with a titration of DRB (0-50 μM) for 4 h and analysed by SDS-PAGE/immunoblot. Levels of protein were detected using antibodies against ZNF350, IRF-1, p53, KAP1 and GAPDH, as indicated. (B) A375 cells were left untreated or treated with 20 mM DRB and harvested at indicated timepoints (0-24 h) post-transfection. Levels of ZNF350 were detected by SDS-PAGE/immunoblot using anti-ZNF350 antibody.

3.2.7 Detection of endogenous IRF-1:ZNF350 complexes in situ

IRF-1 is localised in both the nucleus and the cytoplasm, but being a transcription factor, most of its activity takes place in the nucleus, where it can be translocated by the nuclear localisation signal (NLS) [82; 101]. ZNF350 has been proposed to associate with the nuclear matrix however the localisation of ZNF350 has not been studied in great detail [44]. The functions of some transcription factors are regulated by a variety of mechanisms including transcription, phosphorylation, ubiquitination and protein-protein interactions. Given the critical role of both IRF-1 and ZNF350 in the transcription of genes involved in the DNA damage response, apoptosis and the immune response, the subcellular localization of the proteins could be an important mechanism in regulating their function and ineteraction.

To examine the localization of bound IRF-1 and ZNF350 complexes in situ within a cell, I used proximity ligation assay technology (PLA; Duolink) which enabled me to detect and quantify protein interactions in fixed cells under endogenous protein conditions (Figure 3-15A). This assay works by adding two primary target-specific antibodies, in this case anti-IRF-1 and anti-ZNF350, used together with a pair of species-specific secondary antibodies (PLA probes). These PLA probes contain unique DNA strands that template the hybridization of added oligonucleotides and when in close proximity (<40 nm), the oligonucleotides are ligated by a ligase to form a circular template. This template, still anchored to the antibody, is subsequently amplified and detected using complementary labelled oligonucleotide probes. Detection is performed using fluorescence DNA probes to detect the rolling circle PCR product and interaction events are detected as distinct spots which can be quantified and assigned to specific subcellular locations based on DAPI staining of the nucleus, visualised under a microscope [177; 178] (Figure 3-15A). To initially check the specificity of the proximity ligation assay, I added anti-IRF-1 and anti-ZNF350 antibodies to fixed cells separately (Figure 3-15B, second and third rows, respectively) and then together (Figure 3-15B, bottom row) and found that the spots representing single-molecule protein interactions were only detected when both antibodies were present. This confirms that the assay specifically detects bound IRF-

1 and ZNF350, with most of the interaction events being located within the nucleus (Figure 3-15B, merged, bottom row).

3.2.8 Changes in IRF-1:ZNF350 complex formation

After establishing that IRF-1 and ZNF350 interact *in vivo*, I sought to establish if the cell signals that alter individual protein levels had an effect on the interaction of IRF-1 and ZNF350. A375 cells were treated with different cell stress agents, either IFNγ, etoposide, Poly(I:C), DRB or Trichostatin A, then fixed and analysed using the proximity ligation assay system. The spots representing the single protein interaction event were counted and their localisation either in the nucleus or the cytoplasm was recorded (Figure 3-16). When cells were treated with any of the five agents, the association between IRF-1 and ZNF350 was significantly reduced compared to untreated cells. After treating cells with IFNγ and co-immunoprecipitating the two proteins, there was no clear effect on the interaction, despite IRF-1 levels being greatly increased compared to normal cells (Figure 3-11A and 3-12). When analysed in situ by PLA, IFNγ treatment resulted in a reduction of IRF-1 binding to ZNF350 by about 60% compared to untreated cells (Figure 3-16 and 3-17A). In addition, when examining the distribution of the protein interaction, there does not appear to be a significant alteration compared to untreated cells (Figure 3-17B).

Poly(I:C) treatment significantly reduces IRF-1:ZNF350 binding compared to untreated cells and also results in a reduction in the complex being observed in the nucleus (Figures 3-16 and 3-17). Both IFNγ and poly(I:C) can activate the immune response and induce IRF-1 expression and the effect of both these agents on the association of IRF-1 with ZNF350 is comparable. However, poly(I:C) appears to significantly decrease the presence of the complex in the nucleus compared to untreated cells while there is little change observed with IFNγ.

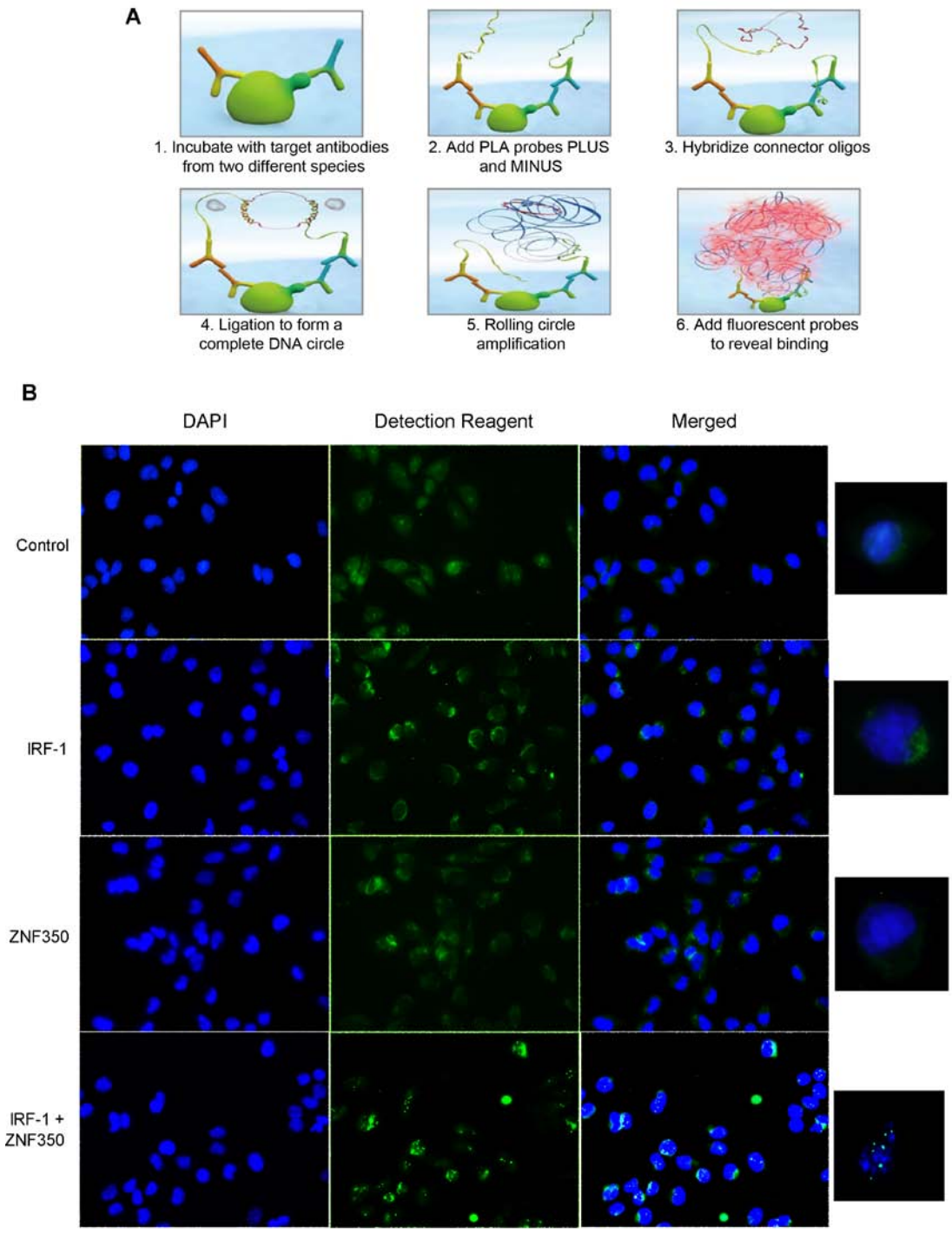


Figure 3-15: Proximity Ligation Assay

(A) Overview of proximity ligation assay methodology (Duolink). (B) PLA was performed using anti-ZNF350 (Abcam 77085) and/or anti-IRF-1 (BD) antibodies as per manufacturer's instruction. Images were captured using an Olympus BX51 microscope with a DAPI filter (blue) and a detection reagent filter (green) then binding was analysed using ImageJ. A close-up of a representative cell is demonstrated on the right.

Poly(I:C) is synthetic double stranded RNA (dsRNA) that can activate the immune response through two distinct pattern recognition receptors. Endosomal poly(I:C) activates TLR3 which induces IL-12 and type I IFN production while cytosolic poly(I:C) activates RIG-1/MDA-5 and stimulates the production of type I IFNs [179; 180]. Interestingly, studies have shown that IRF-1 expressed in response to the combination of poly(I:C) and IFNγ appears to be equivalent to the additive effects of either poly(I:C) or IFNγ alone. This suggests that poly(I:C) and IFNγ stimulate IRF-1 gene expression by two distinct pathways [180]. Furthermore, the promoter region of the *IRF-1* gene contains binding elements for both NF-κB (κB site) and STAT1 (γ-activated site) which overlap creating a composite promoter element to which NF-κB and STAT1 bind in a mutually exclusive and independent manner [181; 182]. dsRNA has been shown to activate NF-κB in several cell types whereas STAT1 is activated in response to IFNγ [183–185]. Despite both poly(I:C) or IFNγ being able to induce the IRF-1-dependent immune response, the two distinct activation pathways for IRF-1 expression may explain the differences in cellular localisation of the ZNF350/IRF-1 complex in response to either poly(I:C) or IFNγ.

The most significant change in the binding dynamics of IRF-1 and ZNF350 is seen after etoposide treatment, with a loss of over 80% of the complex, specifically in the nucleus, compared to that of untreated cells (Figure 3-16 and 3-17). Previous studies showed that when cells were treated with etoposide and analysed by co-immunoprecipitation, a slight decrease in complex was detected, despite the amplification of IRF-1 levels (Figure 3-11B and 3-12). Etoposide is a cytotoxic anticancer drug which acts by forming a ternary complex with DNA and the topoisomerase II enzyme, preventing re-ligation of the DNA strands and causing the strands to break [186]. Following DNA damage, the ATM-signalling pathway is activated resulting, in the up-regulation of IRF-1 mRNA levels and protein synthesis as well as increasing the half-life of the protein [109]. This is distinct from that of the JAK/STAT pathway, which is activated in response to viral agents such as IFNγ and poly(I:C), and activates *IRF-1* gene expression. IRF-1 plays an important role in DNA damage response and has been shown to cooperate with p53 in the G1 checkpoint pathway to induce the cyclin dependent kinase inhibitor $p21^{waf1/cip1}$ in

response to DNA damage [114]. However, despite IRF-1 levels being amplified after etoposide treatment, the interaction with ZNF350 is substantially reduced suggesting either ZNF350 protein levels decrease, which doesn't appear to be the case from previous results (Figure 3-12), or that DNA damage causes the disassembly of the IRF-1/ZNF350 complex in the nucleus. Interestingly, the ZNF350-interacting protein BRCA1 is relocated to the cytoplasm in response to DNA damage via direct binding to p53, with a loss of p53 function resulting in BRCA1 nuclear accumulation [187; 188]. Furthermore, the p53-dependent shuttling determines cellular susceptibility to DNA damage as amplification of cytosolic BRCA1 significantly enhances cancer cell susceptibility to ionizing radiation [188]. This could also be the case with ZNF350 in that its presence in the nucleus following DNA damage may be a hindrance to aspects of the p53-dependent DNA damage response, requiring its ejection from the nucleus and thereby a reduction in IRF-1 complex formation. Gadd45 is a p53-regulated protein that plays a role in the G2-M checkpoint, with its gene expression being induced by DNA damage-inducing agents such as etoposide [189-191]. Conversely, ZNF350 in complex with BRCA1 binds to the *GADD45A* gene downstream of the p53 response element and represses its expression [68]. This, along with the observation that ZNF350 represses the transcription of the p53-inducible $p21^{waf1/cip1}$ DNA damage response gene, highlights some possible reasons for the disruption of the IRF-1/ZNF350 complex in the nucleus [192].

Following treatment with DRB, I found that the levels of IRF-1 and ZNF350 protein levels were significantly reduced in a dose-dependent manner (Figure 3-14). In the proximity ligation assay, there was a significant reduction in ligation signal following the addition of 20μM DRB (Figure 3-16 and 3-17), probably due to a decrease in levels of ZNF350 and IRF-1. Interestingly, it appears more IRF-1/ZNF350 complex is lost from the nucleus than the cytoplasm following DRB treatment, although this is not significant compared to untreated cells (Figure 3-16 and 3-17). This specific reduction in IRF-1 and ZNF350 binding in the nucleus could be due to various factors, including the export of the proteins in response to DRB. DRB has been reported to cause the re-localisation of several proteins such as

p53, poly (ADP-ribose) polymerase (PARP) and poly(A)-binding protein 1 (PABP1). As previously mentioned, p53 protein levels increase following DRB treatment with studies showing it is preferentially accumulated in the cytosol and mitochondria [193]. In the same study, the apoptotic factor Bax, was translocated into the mitochondria and activated by p53, highlighting a possible initiation step in DRB-induced apoptosis [193]. PABP1 is a multifunctional protein found complexed with 3'poly(A) tail of eukaryotic mRNA and is required for poly(A) shortening and translation initiation [194]. PABP1 is found localised in the cytoplasm, however following treatment with DRB it accumulates in the nucleus. This relocation was reversible since removal of DRB restores the cytoplasmic localisation [195]. A similar trafficking pattern has also been reported with the product of the von Hippel-Lindau tumour suppressor gene (*pVHL*), in that inhibition of transcription by DRB resulted in nuclear accumulation of pVHL [196]. Interestingly, although the expression of VHL is mainly cytoplasmic, all of the antitumour properties of VHL tested could be restored by expressing nuclear VHL [197]. These studies would suggest that DRB can induce the re-localization of proteins depending on their requirement for the cell's response to transcription inhibition. Evidently the response to DRB-induced transcription inhibition is complex and the explanation for the reduction of IRF-1 and ZNF350 protein levels and loss of the complex specifically from the nucleus warrants further investigation.

Similar to the effect of all other cell stress treatments on IRF-1 and ZNF350 complex levels, the HDAC inhibitor trichostatin A (TSA) also reduces their interaction in comparison to untreated cells, with a decrease in complex formation detected in the nucleus (Figure 3-16 and 3-17). TSA has been shown to modulate the expression of several genes by inhibiting the activity of HDACs, which causes the acetylation of histones and consequent nucleosome remodelling [198]. Generally, condensed chromatin results in transcriptional repression whereas transcriptionally active genes are found in open chromatin [199]. Inhibition of HDAC activity often allows the heterochromatin to relax and allow for the increase of gene expression. However, global analysis of gene expression has indicated that the inhibition of HDAC activity can result in both the induction and repression of gene expression [200].

TSA has been shown to inhibit IFNγ-induced JAK1 activation and STAT1 tyrosine phosphorylation, which subsequently prevents STAT1 nuclear translocation and STAT1-dependent IRF-1 expression [201]. The fact that TSA negatively regulates STAT1-dependent transcriptional activation suggests that the hyperacetylation of histones negatively regulates STAT-dependent gene activation [201]. A separate study has revealed that TSA can impair the recruitment of IRF-1 to the *IL-12p40* gene promoter. The inhibition of HDACs increased the acetylation at the IL-12p40 locus however, chromatin remodelling, binding of IRF-1 to the promoter and transcriptional activation were inhibited [202]. Although HDAC inhibitors play a fundamental role in regulating gene expression through remodelling of chromatin, they are also known to interfere with the deacetylation of several non-histone proteins including p53 [203] and E2F [204]. It may be possible therefore that TSA can inhibit transcriptional activity through modification of non-histone proteins such as IRF-1. Indeed, IRF-1 has been reported to be acetylated [205] and associate with histone acetyl transferases (HATs) such as p300/CBP to form a multiprotein complex that assembles on the promoter of target genes [206; 207]. This may explain the reduction in interaction of IRF-1 with ZNF350 since post-translational modifications such as acetylation could prevent the two proteins from binding. Alternatively, since ZNF350 is known as a transcriptional repressor, treatment with a transcription-promoting agent such as TSA could result in the dissociation of the IRF-1/ZNF350 repressor complex from to gene promoters. This is supported by a study indicating that ZNF350-mediated repression of HIV-LTR is HDAC2-dependent and when exposed to TSA, repression was lost [56].

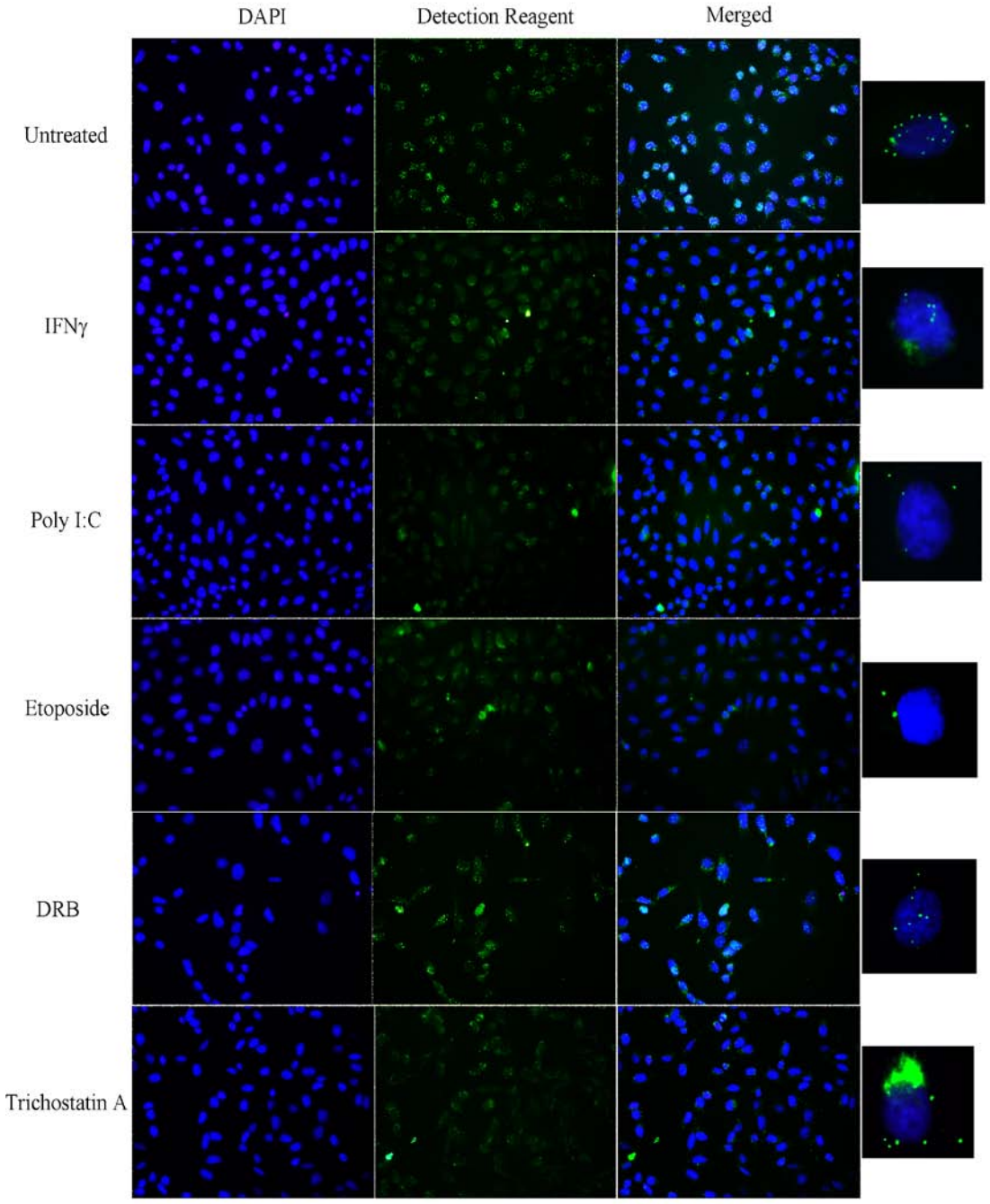


Figure 3-16: ZNF350 and IRF-1 association is reduced following cell stress

A375 cells were treated with IFNγ (200 U/µl, 4h), poly I:C (50 µg/ml, 4h), etoposide (20 µM, 4h), DRB (20 µM, 24h), Trichostatin A (400 nM, 24h) or untreated and a PLA was carried out (Duolink). Images were captured using an Olympus BX51 microscope with a DAPI filter (blue) and a detection reagent filter (green) then binding was analysed using ImageJ. A close-up of a representative cell is demonstrated on the right. Results are representative of 3 independent experiments and at least 6 images of cells were analysed.

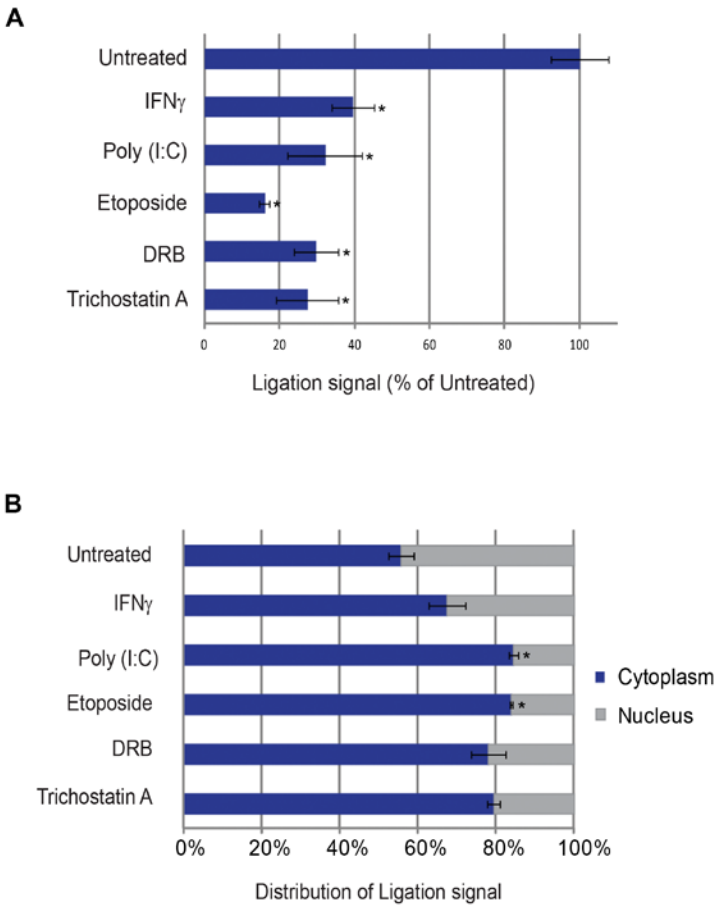


Figure 3-17: Statistical analysis of reduction of protein interaction following cell stress

(A) The occurrence of PLA ligation signal was detected for each condition and calculated as a percentage of untreated ligation signals. (B) The distribution of the ligation signal was detected in relation to the nucleus or cytoplasm condition by counting the number of 'binding spot' then subtracting from the DAPI stain. This leaves spots located outwith the nucleus. Results were calculated as a percentage of untreated ligation signals, and for both A and B, were calculated from 3 independent experiments and at least 6 images were analysed for each. Statistical significance was calculated by t-test and one-way ANOVA. p-value = <0.005.

3.2.9 How do IRF-1 and ZNF350 affect cell growth?

IRF-1 is known to inhibit oncogenesis and several clinical studies have proposed that IRF-1 gene deletion or rearrangement correlates with the development of certain forms of cancer in humans ([110] and references in [208]. The role IRF-1 plays in the regulation of different types of cancer is important for understanding how tumours occur and progress. Cancer metastasis is the most common cause of death among cancer patients [209]. Invasion and metastasis of tumour cells is a multiple process, in which cell motility is accompanied by uncontrolled degradation of basement membrane and components of the extracellular matrix (ECM) [210; 211]. A study into the relationship between ZNF350 mRNA expression and tumour malignancy revealed that in 7 out of 7 cervical cancer cell lines and 9 out of 12 cervical cancer specimens there is significantly reduced amounts of ZNF350 RNA compared to normal tissues. Microarray analysis identified matrix metallopeptidase 9 (*MMP9)* as a potentially downregulated gene target of ZNF350 which was confirmed by knockdown of ZNF350 resulting in elevated expression of *MMP9* in HeLa cells [59]. *MMP9*, also known as gelatinase B (GELB), is capable of degrading type I, IV, V, VII and XI collagens and laminin, suggesting it ultimately regulates cell migration, tumour growth and angiogenesis [212-215]. *MMP9* is overexpressed in many human malignancies including solid tumours and haematological neoplasms [216-218]. ZNF350 has been shown to bind to the putative ZNF350-binding motif on the *MMP9* gene promoter and inhibit *MMP9* gene expression [59].

Interestingly, the *MMP9* promoter also contains several functional enhancer element-binding sites including three AP-1 sites, an Ets site, an SP-1 site, a retinoblastoma element and a non-consensus NFκB binding site [212; 219-223]. TNF-α upregulates the expression of *MMP9* in an NFκB-dependent manner , however this expression is decreased following treatment with IFNβ or IFNγ. Further analysis of the *MMP9* promoter revealed a putative IFN-responsive element (IRE) overlapping the NFκB-binding site to which IFN-induced IRF-1 binds thereby inhibiting NFκB-dependent activation of *MMP9* [148].

Since the *MMP9* promoter contains both a ZNF350 consensus binding site and an IFN-responsive element, I sought to determine the effect of both ZNF350 and IRF-1 has on *MMP9* gene activity. I generated a pGL3-promoter reporter construct of a fragment of the *MMP9* promoter that contains both sites (-1940 to +113) (Figure 3-18A) and found that ectopic expression of either ZNF350 or IRF-1 in HeLa cells significantly reduced the activity of the *MMP9* reporter (Figure 3-18B). Furthermore, when ZNF350 and IRF-1 are co-expressed, their inhibitory effect on *MMP9* reporter activity appears to be synergistic rather than additive (Figure 3-18B). This data suggests that ZNF350 and IRF-1 can cooperate to repress *MMP9* gene expression, a novel mechanism of the complex which would be interesting to explore in other genes containing both ZNF350 and IRF-1 binding sites.

3.2.10 HIV-1 LTR contains IRF-1 and ZNF350 binding sites

3.2.10.1 Introduction

Human immunodeficiency virus (HIV) replication is controlled mainly at the transcriptional level and depends upon a complex interaction between viral and cellular regulatory proteins acting on the viral promoter, the long terminal repeats (LTR) [224-227]. In particular, sequences involved in viral gene expression are contained within the U3 and R regions of the HIV-1 LTR and in the intragenic enhancer in the pol gene [227-229]. The U3 and R regions contain four functional domains: the modulatory and negative regulatory elements spanning nucleotides -454 to nt -104, the enhancer from nt -109 to nt -79, the basal or core promoter from nt -78 to nt -1, and the Tat-responsive element (TAR) from nt +1 to nt +59 [230] (Figure 3-19). Several transcription factors have been found to bind to the HIV-1 LTR with some being ubiquitous while others are modulated in response to cell activation, differentiation, cytokines or mitogens [225]. A unique feature of the HIV-1 LTR is

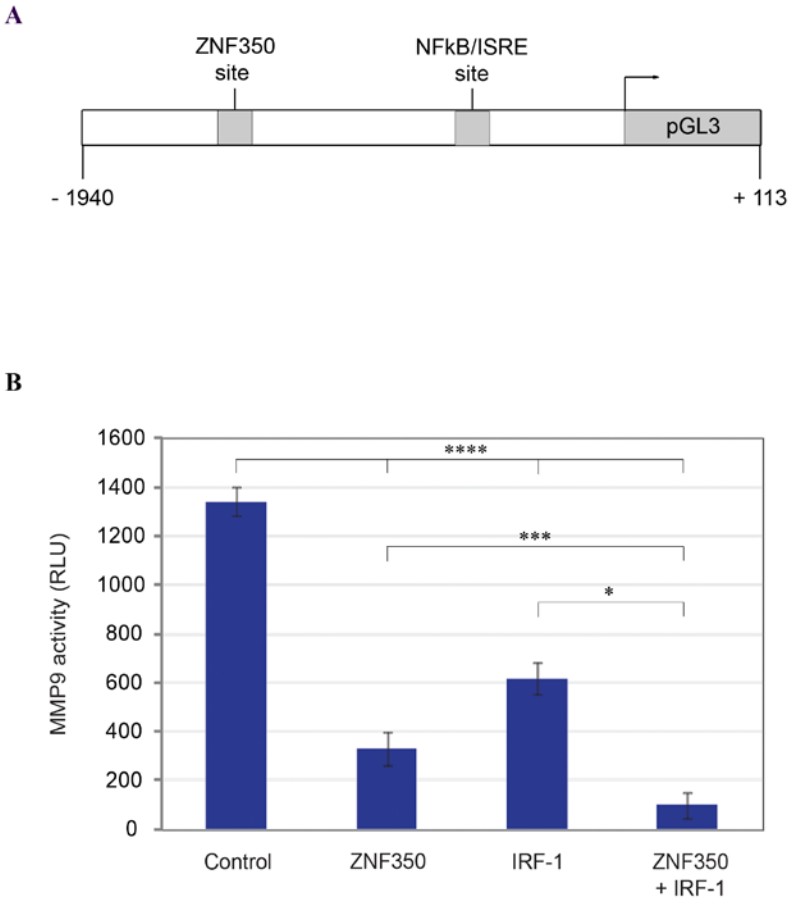


Figure 3-18: ZNF350 and IRF-1 act synergistically to repress *MMP9* gene expression

(A) Schematic of the -1940 to +113 *MMP9*-Luc reporter construct depicting proposed ZNF350 and NFκB/ISRE binding sites. (B) HeLa cells were co-transfected with the *MMP9*-Luc reporter plasmid (100 ng), a control Renilla-Luc plasmid (2 ng), pCMV-ZNF350 (50 ng) and/or pcDNA3-IRF-1 (50 ng), as indicated. DNA amounts were normalised across samples using pcDNA3 empty vector. Post transfection (24 h), the cells were harvested and dual luciferase reporter assays performed. Results were normalised by expressing firefly luciferase/renilla luciferase activity in relative light units (RLU) as the mean +/- S.D. Statistical significance was determined by t-test and one-way ANOVA. ****p-value = <0.0001, ***p-value = <0.001 and *p-value = <0.05.

the presence of transcriptional regulatory elements found downstream of the transcription start site. These sites include three AP-1 sites, an NF-AT site, two SP1 sites and a sequence that is homologous to the ISRE spanning nt +200 to nt +217 [229; 231-233]. These sequences have been found to be strongly conserved in different HIV-1 isolates [230]. Studies of mutant viruses in which these motifs are altered have revealed that the regulatory sites are critical for viral transcription and replication, in particular the region between nt +200 and +217 containing the ISRE. The ISRE is located in a DNAse I-hypersensitivity site [234] and has been shown to be a binding site for members of the IRF family, in particular IRF-1 and IRF-2 [229; 235]. Studies have shown that IRF-1 is induced by HIV-1 infection at early time points before transcripts of the HIV-1 transactivator protein, Tat, are detected and is able, *per se*, to activate transcription from the LTR [235]. Tat is produced early after infection and before viral integration [236], and is critical for productive viral replication [237; 238]. It has been shown to modulate viral gene expression by increasing the rate of transcription initiation, elongation and translation of transactivation response element (TAR)-containing mRNAs [239-241]. In the presence of low doses of Tat, IRF-1 is reported to increase Tat-mediated HIV-1 transactivation by direct physical interaction [235]. In addition, both Tat and IRF-1 have been shown to functionally interact with general transcription factors such as TFIIB [242; 243] and co-activators or adaptors including the histone acetyltransferases p300/CBP and pCAF [207; 244]. Interestingly, this positive cooperation is blocked by IRF-8, a physiological repressor of IRF-1 activity, which inhibits Tat-mediated LTR transcription and viral replication *in vivo* [235].

Further studies into the role of IRF-1 in HIV-1 LTR transcriptional regulation revealed that IRF-1 binds to the enhancer κB sites in combination with NFκB p50/p65 heterodimer and is required for full NFκB transcriptional activity at the HIV-1 LTR enhancer. When IRF-1 is silenced by small interfering RNA (siRNA), NFκB-mediated transcriptional activation of the HIV-1 LTR after infection of T cells was impaired [245].

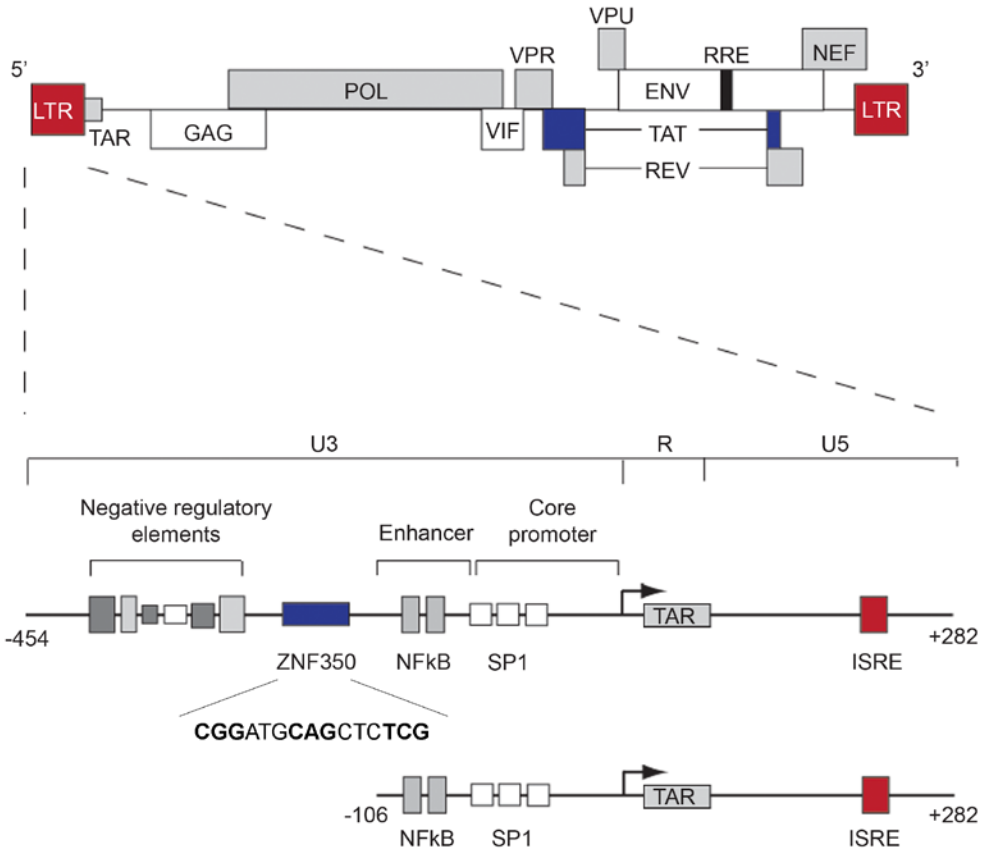


Figure 3-19: HIV-1 LTR

Schematic depicting the domains of HIV-1 (upper panel) and the 5'LTR (lower panel), containing the proposed ZNF350 binding site.

A genome-scale RNAi screen identified the ZNF350-interacting protein, BRCA1, as a host factor for HIV-1 replication [246]. This, along with the finding that another ZNF350-binding protein KAP1, can inhibit HIV-1 replication by suppressing integrase acetylation during integration [247], led to Nishitsuji *et al.* [56] to investigate the possible role of ZNF350 in HIV-1 replication. They found that ZNF350 can bind to a site located in the HIV-1 LTR (between nt -145 to -126) and suppress its promoter activity. In addition, they found that this repression was not affected by mutations in the NFκB or SP1 cis-regulatory elements within the HIV-1 LTR or by the absence of Tat. Furthermore, the ZNF350 repression of HIV-1 LTR is dependent on KAP1 and HDAC2, however not specifically the KAP1-binding histone deacetylase SETDB1 or HP1 [56].

3.2.10.2 Results

The presence of both an IRF-1-binding ISRE and a potential ZNF350 binding site on the HIV-1 LTR led me to question whether the two proteins may act together to repress HIV-1 replication. Upon closer inspection of the HIV-1 LTR region potentially responsible for ZNF350 binding (-145 to -126), reported by Nishitsuji *et al.* [56], I found a sequence quite similar to that of the putative ZNF350 consensus motif (GGGxxxCAGxxxTTT) (Figure 3-19). Two HIV-1 LTR reporter constructs used in the ZNF350 repression study were kindly gifted by the authors [56] to investigate the effect of both IRF-1 and ZNF350 on the activity of the HIV-1 LTR. Nishitsuji *et al.*[56] co-transfected HEK 293T cells with 200 ng of Flag-ZNF350 and HIV-1 LTR-driven luciferase reporter plasmid and determined the activity of full-length (-454 to +282) or truncated (-106 to +282) HIV-1 LTR 48 h post-transfection. I decided to replicate these conditions with untagged ZNF350 as well as IRF-1 to determine its affect on HIV-1 LTR activity alone and in combination with one another (Figure 3-20). Ectopic expression of ZNF350 significantly repressed full length HIV-1 LTR activity but not the truncated mutant activity, supporting the previous results indicating the presence of a ZNF350 binding motif located between -145 to -126 (Figure 3-19A and B). Curiously, expression of IRF-1 also significantly

repressed the activity of both truncated and full-length HIV-1 LTR, to a greater degree to that of ZNF350 expression. This result contradicts previous findings that IRF-1 can actually activate HIV-1 LTR [245]. When IRF-1 and ZNF350 are co-expressed, the effect on HIV-1 LTR activity is comparable to that seen with IRF-1 expression alone, indicating that ZNF350 has little effect. This would also suggest that these two transcription factors do not act synergistically or additively at the HIV-1 LTR promoter to repress activation (Figure 3-20A). In addition, the loss of the ZNF350 binding motif does not affect the repression of the HIV-1 LTR by IRF-1, again suggesting ZNF350 is not required for IRF-1-dependent repression (Figure 3-20B).

Since my finding that IRF-1 decreases HIV-1 LTR activity contradicts previous studies, I decided to investigate the effect of different conditions on LTR activity. I co-transfected HEK 293T cells with full-length or mutant HIV-1 LTR-driven luciferase reporter plasmids and different concentrations (50-200 ng) of ZNF350 and IRF-1 and measured the luciferase activity 48 h after transfection. ZNF350 gradually repressed full length HIV-1 LTR activity in a dose-dependent manner (Figure 3-21A), however the concentration of ZNF350 did not appear to reduce the activity of the HIV-1 LTR mutant and appears to in fact slightly increase activity (Figure 3-21B). IRF-1 also represses both full-length and mutant HIV-1 LTR quite profoundly (Figure 3-21A and B, respectively) but higher concentrations of IRF-1 does not have any greater effect on HIV-1 LTR expression and repression appears to be optimum at 50 ng. Additionally, when IRF-1 and ZNF350 are co-expressed, the effect on HIV-1 LTR activity is comparable to that seen with IRF-1 expression alone with both full-length and truncated HIV-1 LTR (Figure 3-21A and B, respectively). Next, I looked at the activity of the LTR at different time points post-transfection of ZNF350 or IRF-1 (100 ng; Figure 3-21C and D). As the time between transfection of ZNF350 and measurement of luciferase activity increases, the affect on full-length HIV-1 LTR activity appears to be slightly reduced (Figure 3-21C). Interestingly, 24 to 32 h after transfection of ZNF350 there is a clear repression of truncated HIV-1 LTR activity however 48 to 56 h post-transfection, this inhibition is lost (Figure 3-21D). On the other hand, the length of time between IRF-1 transfection and the

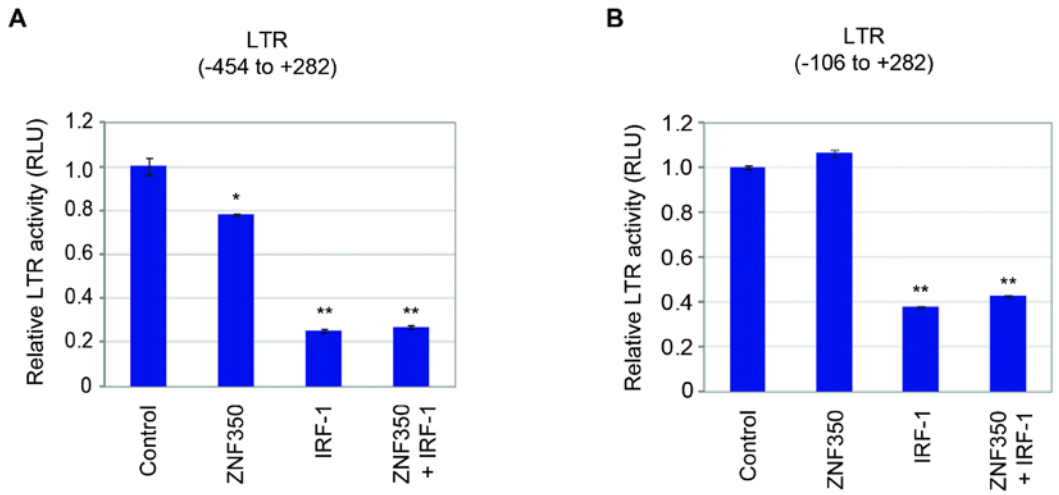


Figure 3-20: ZNF350 and IRF-1 repress the activity of HIV-1 LTR

HEK 293T cells were co-transfected with a HIV-1-LTR (-454 to +282)-Luc reporter plasmid (1 ng per well), a control Renilla-Luc plasmid (1 ng per well), pCMV-ZNF350 (100 ng) and/or pcDNA3-IRF-1 (100 ng), as indicated. DNA amounts were normalised across samples using pcDNA3 empty vector. Post transfection (48 h), the cells were harvested and dual luciferase reporter assays performed. Results were normalised by expressing firefly luciferase/renilla luciferase activity in relative light units (RLU) as the mean +/- S.D. *p-value = <0.01 or **p-value = <0.0001, determined by t-test. (B) As in (A), except that a HIV-1-LTR (-106 to +282)-Luc reporter plasmid was used instead of the (-454 to +282) reporter construct.

measurement of HIV-1 LTR gene expression does not appear to affect the inhibitory affect of IRF-1 on activity of both full-length and mutant LTR (Figure 3-21C and D).

The data presented here is representative of at least six independent experiments, however the result that IRF-1 significantly reduces the activity of the HIV-1 LTR contradicts previous findings that IRF-1 can actually activate LTR expression. The luciferase assay was performed under exactly the same conditions as Nishitsuji et al., as far as I can tell, including cell type, plasmid concentration, transfection and assay type, with the comparable repression of HIV-1 LTR by ZNF350 supporting this. Previous studies showing IRF-1 can activate HIV-1 LTR expression have been performed in various cell types including HEK 293T cells, with similar concentrations of IRF-1 (10-750 ng) [230; 235; 245; 248; 249]. However, these studies used a reporter gene encoding chloramphenicol acetyl transferase (CAT) where activity is measured from expressed proteins, using 14C-labelled chloramphenicol as the substrate [250]. One possible way to explain the apparent repression of HIV-1 LTR by IRF-1 comes from a study showing that mutations in both NFκB enhancer sites impaired the activation of the LTR by IRF-1 and actually resulted in a decrease in expression compared to wt LTR [245]. However, to rule out the possibility that there may be mutations, I sequenced the HIV-1 LTR reporter constructs and did not find any discrepancies. Studies have indicated that deletion or mutation in the ISRE sequence leads to a general defect in the viral promoter activity in the presence of Tat [249; 251]. Furthermore, the replacement of the LTR-ISRE with another functional ISRE from the ISG54 gene leads to a decrease in both LTR promoter activity and proviral DNA gene expression [249]. It is possible that the +200 to +217 element in the HIV-1 LTR is not functioning as an ISRE in the regulation of HIV-1 gene expression [230]. The mechanism by which the ISRE element may alter HIV-1 LTR promoter activity is still uncertain, however it may be that the sequence could act cooperatively with other elements to position the nucleosomal structure around the LTR to regulate promoter activity in the context of chromatin [231; 249]. The reason for the discrepancies between the results presented here and previous findings are as yet unidentified however, it would be of great interest if discovered since such a significant repression of HIV-1 is observed.

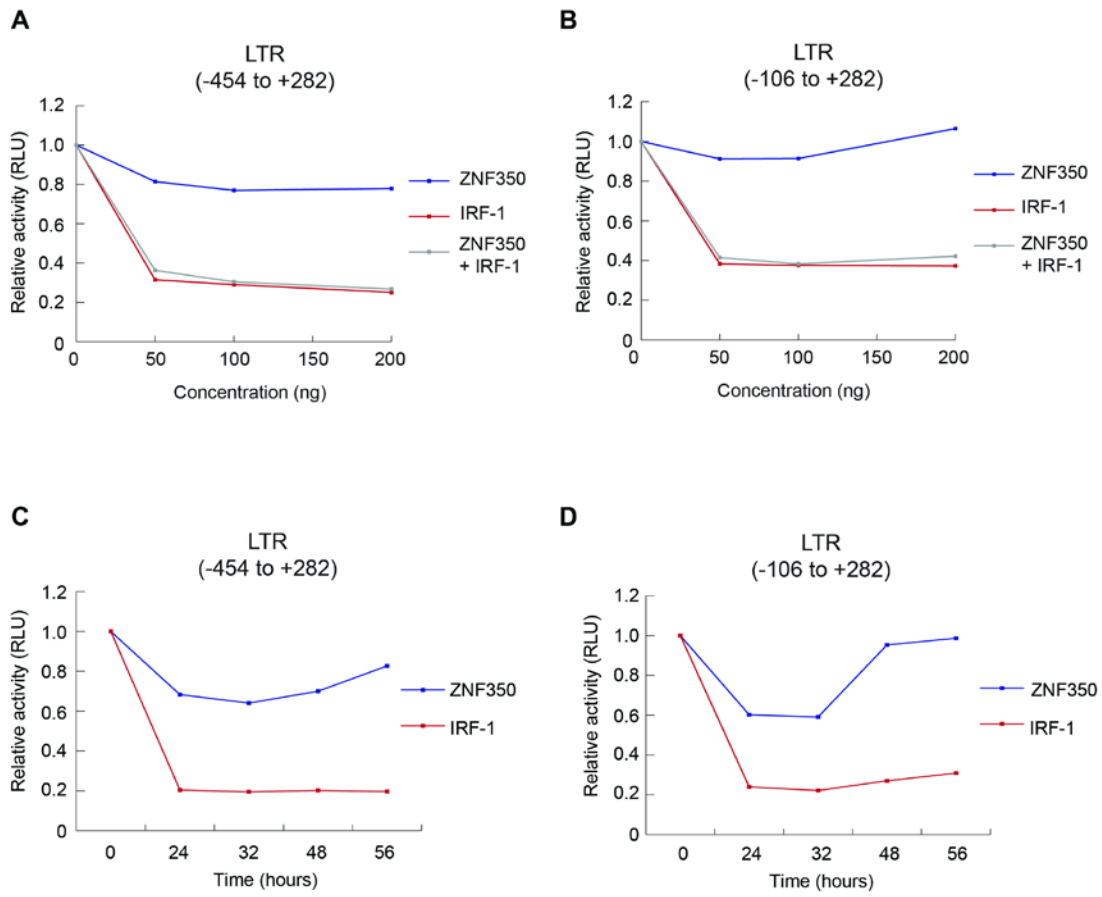


Figure 3-21: Dose-dependent and timecourse effects of ZNF350 and IRF-1 on HIV-1 LTR activity

HEK 293T cells were co-transfected with (A) HIV-1-LTR (-454 to +282)-Luc reporter plasmid or (B) HIV-1-LTR (-106 to +282)-Luc reporter plasmid (both 1 ng per well), along with a control Renilla-Luc plasmid (1 ng per well), and titrations of pCMV-ZNF350 (0-200 ng) and/or pcDNA3-IRF-1 (0-200 ng), as indicated. DNA amounts were normalised across samples using pcDNA3 empty vector. Post transfection (48 h), the cells were harvested and dual luciferase reporter assays performed. Results were normalised by expressing firefly luciferase/renilla luciferase activity in relatve light units (RLU) as the mean +/- S.D. (C) Dual luciferase reporter assays performed as in (A) with 100 ng HIV-1-LTR (-454 to +282)-Luc reporter plasmid and cells harvested 0-56 h post transfection, as indicated. (D) As in (C) except with HIV-1-LTR (-106 to +282)-Luc reporter plasmid instead of HIV-1-LTR (-454 to +282) reporter construct.

3.3 Discussion

The C-terminal Mf1 domain is a key regulatory domain involved in controlling IRF-1-dependent effects on gene expression. It contains several sub-domains including: (i) a region that negatively regulates IFNβ activation [98], (ii) a degradation signal [101], (iii) a motif that represses *Cdk2* expression which is also involved in IRF-1-dependent cell growth suppression [98], and (iv) a region that appears to regulate the stability of IRF-1 [102]. Despite structure-function analysis revealing the Mf1 domain to be required for several IRF-1 activities, to date only members of the Hsp70 family have been identified as Mf1 binding proteins [161]. Following data from a yeast-2 hybrid screen and peptide phage display suggesting the KRAB-containing zinc finger protein ZNF350 may interact with the C-terminus of IRF-1, I have been able to validate this interaction using a combination of *in vitro* and cell experiments. I showed that ZNF350 directly interacts with IRF-1 in a cellular environment by co-immunoprecipitating the two proteins (Figure 3-2B and C) and by detecting the endogenous complexes in situ (Figure 3-16 and 3-17). Further analysis using peptide pulldowns with IRF-1 peptides revealed that ZNF350 specifically interacts with the IRF-1 nuclear localisation signal and the Mf1 domain (Figure 3-9). Fine mapping of the interaction between ZNF350 and the C-terminus of IRF-1 demonstrated that a co-signature LXXLL motif, along with several proline and threonine residues, was an essential component of the interaction (Figure 3-10B).

The LXXLL motif within the IRF-1 C-terminus appears to be critical for the role of the MF1domain in regulating both the biological activity [98] and ubiquitin proteasome degradation of IRF-1 [101]. It has been show that mutation of this motif significantly impairs IRF-1's growth suppressor activity and IRF-1-dependent *Cdk2* repression as well as rendering the IRF-1 protein resistant to proteasomal degradation [98; 101]. LXXLL motifs are present in many transcription factors and mediate protein-protein interactions that often function in activating or repressing transcription [100]. The fact that ZNF350 binds to the LXXLL provides compelling evidence for a role for it in the regulation of IRF-1-mediated gene expression

It is therefore possible that ZNF350 binding to the LXXLL motif of IRF-1 may be required for IRF-1 to repress gene expression as the model of ZNF350 action requires it to bind to DNA. Finding IRF-1 and ZNF350 binding elements within the same gene regulatory domains will be an important step in understanding the significance of the interaction.

IRF-1 protein is present in cells at very low levels and its expression rapidly induced by stimuli such as virus, dsRNA, IFNs, tumour necrosis factor (TNF), IL-1, IL-6, prolactin, retinoic acid and leukemia inhibitory factor (LIF) [106; 164]. Very little is known about the expression of the ZNF350 protein so I sought to determine the effect of several IRF-1 stimulants, namely IFNγ and the DNA damage mimetic etoposide. While the expression of ZNF350 in cell lysates was only slightly induced compared to IRF-1 by these agents (Figures 3-10 and 3-11), the interaction between IRF-1 and ZNF350 was significantly reduced when analysed using a proximity ligation assay (Figure 3-16 and 3-17). In addition, after treatment with either poly(I:C) or etoposide, about 80% of the total cell population of ZNF350/IRF-1 complex was present in the cytoplasm, compared to just over 50% in untreated cells. This would indicate that the complex is disrupted in the nucleus following treatment with etoposide or poly(I:C) (Figure 3-16 and 3-17). Since IRF-1 enhances the transactivation of IFN- and DNA damage-inducible genes and the primary activity of ZNF350 is to repress gene expression, it is interesting that their interaction is lost under these conditions. UV light is a potent stimulus for the release of inflammatory cytokines and interestingly, can induce almost all cytokines. However, UV has been found to suppress IFNγ-induced expression of the cytokine IL-7 by reducing IRF-1 mRNA levels [252]. I found that upon UV irradiation, both IRF-1 and ZNF350 proteins levels are decreased in a dose-dependent manner and as a result their incidence of interaction was also reduced (Figure 3-13).

IRF-1 and ZNF350 both play a role in the regulation of gene expression, with IRF-1 mostly enhancing expression in response to various stimuli and ZNF350 repressing expression under as yet unknown conditions. Transcriptional control is mediated by various regulatory components including RNA Pol II, activators, and repressors;

HATs, HMTs and HDACs. Generally, condensed chromatin results in transcriptional repression whereas transcriptionally active genes are found in open chromatin [199; 253]. Transcriptional activators bind to DNA and recruit HATs which acetylate conserved lysine residues on histone N-terminal tails removing the positive charge and decreasing the interaction with the negatively charged phosphate groups of DNA. This results in a more open, relaxed chromatin structure, enabling the initiation of transcription by RNA Pol II [253; 254]. On the other hand, transcriptional repressors bind to gene promoters and recruit HDACs, which remove the acetyl groups from histones, increasing their interaction with DNA and ultimately condensing chromatin rendering it mostly inactive [23; 253; 255].

In order for transcription to begin, RNA Pol II C-terminal domain must be phosphorylated to enable a conformational change to release it from the promoter. DRB is a potent inhibitor of the kinases responsible for this phosphorylation [166; 167] and has been shown to ultimately inhibit transcription *in vivo* [170]. Both IRF-1 and ZNF350 protein levels are reduced in response to DRB, in a dose-dependent manner and the interaction between the proteins is significantly inhibited with a slight re-localisation to the cytoplasm (Figures 3-14, 3-16 and 3-17). It is possible that the reduction in levels of both proteins and consequent decrease in binding is due to the inhibition of their gene expression by DRB, and this could be determined by further analysis of gene activation and mRNA levels within cells.

Another transcriptional regulator is the HDAC inhibitor TSA which has a similar effect on IRF-1/ZNF350 complex levels and loss of nuclear complex formation as DRB (Figure 3-16 and 3-17). TSA however, inhibits deacetylation of histones therefore allowing for hyperacetylation, the relaxation of heterochromatin and subsequent increased gene expression [198]. However, global analysis of gene expression has indicated that the inhibition of HDAC activity can result in both the induction and repression of gene expression [200] and TSA can in fact result in the post-translation modification of non-histone proteins [205]. Thus, it is possible that the acetylation (or inhibition of deacetylation) of IRF-1 may disrupt the interaction with ZNF350. The data presented here, shows that the IRF-1/ZNF350 can be disrupted by a variety of cell treatments which can either induce IRF-1 expression or

repress it. This indicates that the interaction between the two proteins may be quite sensitive and transient and will warrant further investigation to elucidate how complex formation is modulated.

Enhanced expression of *MMP9* correlates with invasion during tumour progression (references in [148]). Interferons have been shown to inhibit *MMP9* activation in response to TNF-α, which activates the *MMP9* gene through NFkB (references in [148]). It was subsequently discovered that IRF-1 is induced by IFNs and inhibits the NFkB-dependent activation of *MMP9* [148]. In a separate study, ZNF350 was show to bind to a putative ZNF350-binding site on the *MMP9* promoter and inhibit its activity [59]. The data presented here supports the observation that IRF-1 and ZNF350 can inhibit *MMP9* activity separately. I have then expanded this knowledge to demonstrate that the two proteins work synergistically at the *MMP9* promoter to repress its expression (Figure 3-18). This finding is of particular interest since understanding the molecular basis for *MMP9* inhibition may provide tools to control cell invasion.

Following the identification of the synergistic activity displayed by IRF-1 and ZNF350 during the repression of *MMP9*, I hoped to see the same repression mechanism with HIV-1 LTR. The HIV-1 LTR, similar to the *MMP9* promoter, possesses both a ZNF350 consensus motif and an ISRE [56; 251]. While ZNF350 has been shown to repress HIV-1 LTR activity, studies have shown that IRF-1 is induced by HIV-1 infection and is able to activate transcription of the LTR [235]. The data presented here verifies the repression of LTR by ZNF350, however contrary to the IRF-1 study, my data actually shows that IRF-1 can significantly repress HIV-1 LTR activity. In addition, this repression appears to be independent of ZNF350 activity due to the comparable repression following loss of ZNF350 binding (Figure 3-20 and 3-21). The mechanism by which IRF-1 can repress HIV-1 LTR activity remains to be elucidated but will be of great interest since it may provide insight into how to control HIV-1 replication.

The results presented in this chapter clearly demonstrate a direct interaction between ZNF350 and IRF-1 and begin to determine how these two proteins interact in vivo.

The ZNF350/IRF-1 complex does not appear to form in response to stress and rather is present in the basal state of cells. In this case, the complex could be important in controlling homeostasis and further studies will hopefully reveal the true extent of the relationship between IRF-1 and ZNF350 within a cellular environment. To expand on this data, Chapter 4 examines the formation of the complex on chromatin and identifies novel genes possessing potential binding sites for both proteins.

Chapter 4: Identification of novel gene targets of ZNF350 and IRF-1

4.1 Introduction

4.1.1 Transcriptional Control

Gene specific repression of transcription plays a central role in gene regulation. Transcriptional control is mediated by a wide variety of regulatory components involving a complex and diverse range of protein-DNA and protein-protein interactions.

Eukaryotic gene transcription is controlled by epigenetic changes and chromatin reorganization, which are mediated by sequence-specific transcription factor binding to proximal or distal regulatory elements [256; 257]. While transcription factors can bind DNA directly, they are unable to modify histones, which is a key step in chromatin reorganization. However, they act as a scaffold on to which co-activators or repressors bind and either directly modify histone tails or recruit additional proteins such as HATs, HDACs or HMTs [258]. The result of this network of interactions is the formation of transcriptional complexes on DNA that can consist of up to 50 different proteins [259], which act together to determine the activation or repression of the gene and ultimately alter cell function [257].

Many repressors are known to act via several mechanisms and the mechanism used by a particular repressor is often promoter dependent. For example, the mammalian cell cycle regulator and tumour suppressor protein Retinoblastoma (Rb) uses multiple mechanisms to repress transcription and represses different promoters using different combinations of mechanisms ([260-262] and references in [2]). Gene-

specific repressors can regulate transcription by either indirectly or directly binding to DNA close to the promoter (short-range) or a considerable distance up or downstream (long-range). 'Long-range' repression can occur when the repressor protein affects chromatin structure at or near the promoter [2]. Another suggested method of long-range repression involves the repressor protein being able to make contact with activators and/or components of the PIC by 'looping' of the chromatin [13]. Conversely, 'short-range' repression takes place when the repressor acts locally at the promoter to affect transcription. Interestingly, a repressor protein may cause both long and short range repression depending on its binding site and mechanism of action [2; 11].

ZNF350 has been shown to bind directly to several different genes including *GADD45A*, *ANG1*, *p21*, *MMP9* and *HIV-1 LTR* and repress their activity [45; 52; 55; 56; 59]. However in each of these cases, ZNF350 binds to various regions of the gene and in combination with different proteins. For example, a complex of BRCA1, ZNF350 and HDACs bind to intron 3 of GADD45A to suppress gene expression [45] while a KAP1/LCoR/ZNF350 complex can also bind to this region separately and repress expression [51]. To date, the gene targets of ZNF350 have roles in various different cellular functions such as DNA damage response, cell cycle regulation, angiogenesis and metastasis. However, the role of ZNF350 in determining gene regulation and cellular function is largely unexplored.

IRF-1 was first shown to bind to DNA sequences upstream of the *IFN-β* gene [263] with further single-gene studies identifying a number of targets of IRF-1. These include MHC class II transactivator CIITA [264], IL-15 [265], IL-12 [266], LMP-2 and TAP-1 [267], which are involved in the immune response; and p21 and caspase-1 which are targeted by IRF-1 in DNA damaged cells. In addition, IRF-1 has been shown to bind to genes involved in apoptosis such as caspase-8, PUMA, TRAIL and XAF1 [268-271]. A more recent study utilising ChIP-chip analysis identified a number of novel IRF-1 targets in IFN-γ treated breast cancer cells such as BRIP1, which is involved in the DNA damage pathway [272]. In addition, ChIP-seq carried out in human primary monocytes identified 52 IRF-1-bound regions associated with

genes involved in the immune response such as ERAP1, GBP1, TLR3 and host-virus interaction genes including SP110, TAP1 and TRIM5 [273]. A further ChIP-seq study in breast cancer cells has revealed that IRF-1 binds to a large number of genomic regions associated with apoptosis in both unstimulated and IFN-γ stimulated cells [274].

This chapter aims to describe the regulatory targets of both ZNF350 and IRF-1 within the genome in order to better understand the relationship between the two proteins. By using a combination of cell-based assays and bioinformatics studies I have identified genomic regions that harbour binding sites for both ZNF350 and IRF-1.

4.2 Results

4.2.1 ZNF350 tightly associates with chromatin

Data from Chapter 3 and previous studies revealed that ZNF350 is localised in the nucleus and can directly bind to DNA. To confirm that ZNF350 binds directly to DNA *in vivo* I set up an assay to specifically look at chromatin-bound ZNF350 using a 10-50% isokinetic sucrose gradient. Sucrose density gradient ultracentrifugation is a technique used to analyse chromatin structures where the rate of sedimentation through the gradient is determined by the mass and hydrodynamic shape of the chromatin fibres [275; 276]. In this way, large, bulky structures such as euchromatin sediments towards the bottom while looser, open chromatin will remain close to the top of the gradient (Figure 4-1A). The nuclei from A375 cells was isolated and chromosomal DNA was briefly digested using micrococcal nuclease, After RNase treatment of the samples to remove any RNA, the nuclei were lysed and the lysate was fractionated using sucrose density gradient centrifugation. Analysis of the UV trace and examination of the DNA content of each fraction by agarose gel analysis following phenol/chloroform extraction revealed that the chromosomal DNA, and consequently the proteins associated with it was mainly present in fractions 6-8. Alternatively, earlier fractions (1-3) contain nuclear proteins that are not associated with chromatin while fractions 4 and 5 contain proteins only loosely associated with the chromatin (Figure 4-1B). Protein analysis of the samples by ethanol precipitation followed by SDS-PAGE shows that ZNF350 is present in later fractions (7-10) revealing it is tightly associated with chromatin, supported by the correlation with Histone H3 protein in the same fractions (Figure 4-1C). The presence of IRF-1 in the chromatin fractions was also analysed but the results were often variable and it was difficult to come to any definitive conclusions on IRF-1's association with the chromatin.

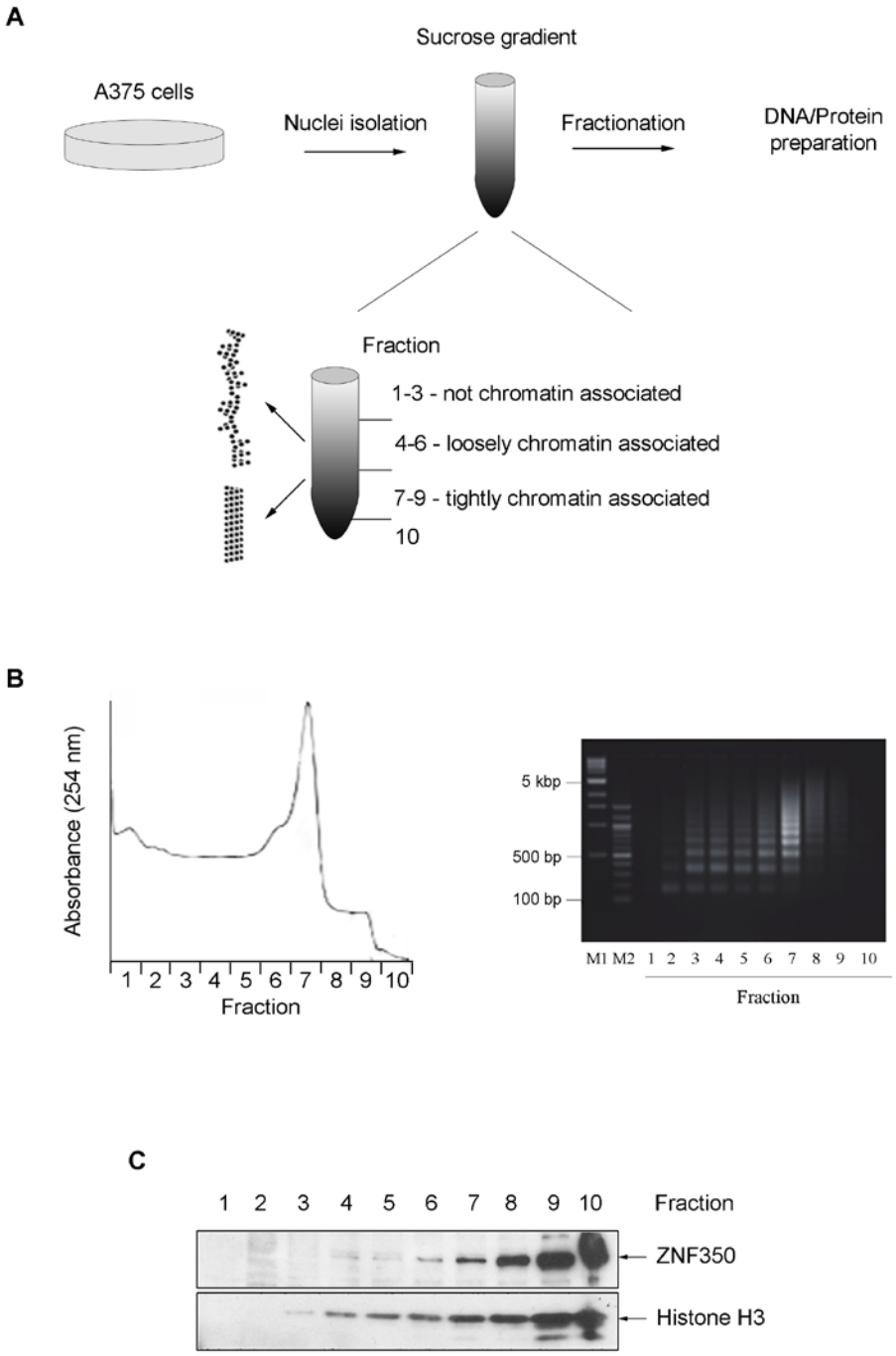

A
Sucrose gradient
A375 cells
Nuclei isolation
Fractionation
DNA/Protein preparation
Fraction
1-3 - not chromatin associated
4-6 - loosely chromatin associated
7-9 - tightly chromatin associated
10
B
Absorbance (254 nm)
1 2 3 4 5 6 7 8 9 10
Fraction
5 kbp
500 bp
100 bp
M1 M2 1 2 3 4 5 6 7 8 9 10
Fraction
C
1 2 3 4 5 6 7 8 9 10 Fraction
ZNF350
Histone H3

Figure 4-1: ZNF350 tightly associates with chromatin

(A) A375 cells are grown in a tissue culture dish, treated as indicated and harvested at around 90% confluency. Nuclei were isolated, RNA was removed in an RNase step and chromosomal DNA partially digested using micrococcal nuclease. Following digestions, cells were lysed and the lysate is separated using a 10-50% sucrose gradient and ten 0.5 ml fractions collected using a fraction collector. During fraction collection a UV profile of the fractions was recorded. (B) Typical UV trace of a fractionation experiment (left panel) and agarose gel representing DNA content of the ten fractions obtained (right panel). DNA was prepared by phenol/chloroform extraction followed by ethanol precipitation (experiment performed by Vivien Landre). (C) Protein content of nuclei was analysed by SDS-PAGE/immunoblot with anti-ZNF350 mAb, Histone H3 mAb or anti-IRF-1 mAB as indicated.

4.2.2 Chromatin immunoprecipitation

From the confirmation that ZNF350 tightly associates with chromatin and previous published data describing the regulation of various genes by ZNF350 and IRF-1 individually, I sought to expand this knowledge and identify novel binding sites for both ZNF350 and IRF-1 within the genome.

Genome-wide studies of protein-DNA interactions and epigenetic marks are essential for a full understanding of transcriptional regulation. An accurate map of binding sites for transcription factors, transcriptional machinery and other DNA binding proteins such as repressors and activators is essential for understanding the networks of gene regulation that determine various biological processes [277]. The main tool for investigating these mechanisms is chromatin immunoprecipitation (ChIP), which is a technique for establishing protein-DNA binding *in vivo* [278]. Generally in ChIP, antibodies are used to select specific proteins, which enriches for DNA fragments that are bound to these proteins. In the past, microarrays allowed the DNA fragments obtained from ChIP to be identified by hybridization to a microarray (ChIP-chip), thereby showing a genome-scale view of DNA-protein interaction [279; 280]. More recently, the rapid technological advances in high-throughput sequencing, also known as next-generation sequencing (NGS), has provided the ability to sequence tens or hundreds of millions of short DNA fragments in a single run [281-283]. This has allowed for the application of NGS to be used to sequence DNA fragments of interest obtained from ChIP (ChIP-seq) [284]. In ChIP-seq, the DNA fragments are directly sequenced instead of being hybridized on a microarray. This allows for higher resolution, fewer artefacts, greater coverage and a larger dynamic range than ChIP-chip and provides considerably improved data [285]. ChIP-seq has been used widely for many transcription factors, histone modifications, chromatin modifying complexes and other chromatin-associated proteins in a wide variety of organisms [286].

In order to perform ChIP-seq with ZNF350 and IRF-1, I started out by optimizing ChIP to ensure maximum accuracy and success with DNA sequencing. An outline of the ChIP protocol is given in Figure 4-2. Briefly, cells are treated with a chemical agent, such as formaldehyde, to cross-link proteins covalently to DNA. This is followed by cell disruption and sonication to shear the chromatin to a target size of 200-800 bp. The protein of interest with its bound DNA is then enriched comparative to the starting chromatin by purification with an antibody specific for the factor. After this immuno-enrichment, the cross-links are reversed and the enriched DNA is purified and prepared for analysis [286].

4.2.2.1 Chromatin extraction form cells

The cell type, cell number and lysis buffer used for ChIP can have a great impact on the efficiency of protein extraction and DNA shearing. For this reason, I started out by analysing different cell types and nuclei lysis buffers in order to obtain the highest abundance of protein extraction. Using five different cell types: A375, HeLa, H1299 and the human glioblastoma cell lines U251 and SF539, I initially extracted cellular proteins by lysing with a gentle lysis buffer. After collecting the insoluble nuclear components in the pellet, I performed a further lysis step using four different lysis buffers to analyse which resulted in the most protein extraction (Figure 4-3). I ascertained that by using buffer 3 (RIPA buffer) to lyse A375 cells, I was able to extract the highest levels of ZNF350 and IRF-1 protein (Figure 4-3A) compared to other lysis buffers and cell types.

Following issues with the shearing of DNA, I tested two different sonication devices, a Soniprep tip sonicator (Sanyo) and a Bioruptor (Diagenode), using different numbers of sonication pulses. Despite analysis by agarose gel indicating the DNA was of a particular size, often this did not translate to the sample as a whole, as most of the DNA remained largely unsheared following analysis by the sequencing facility

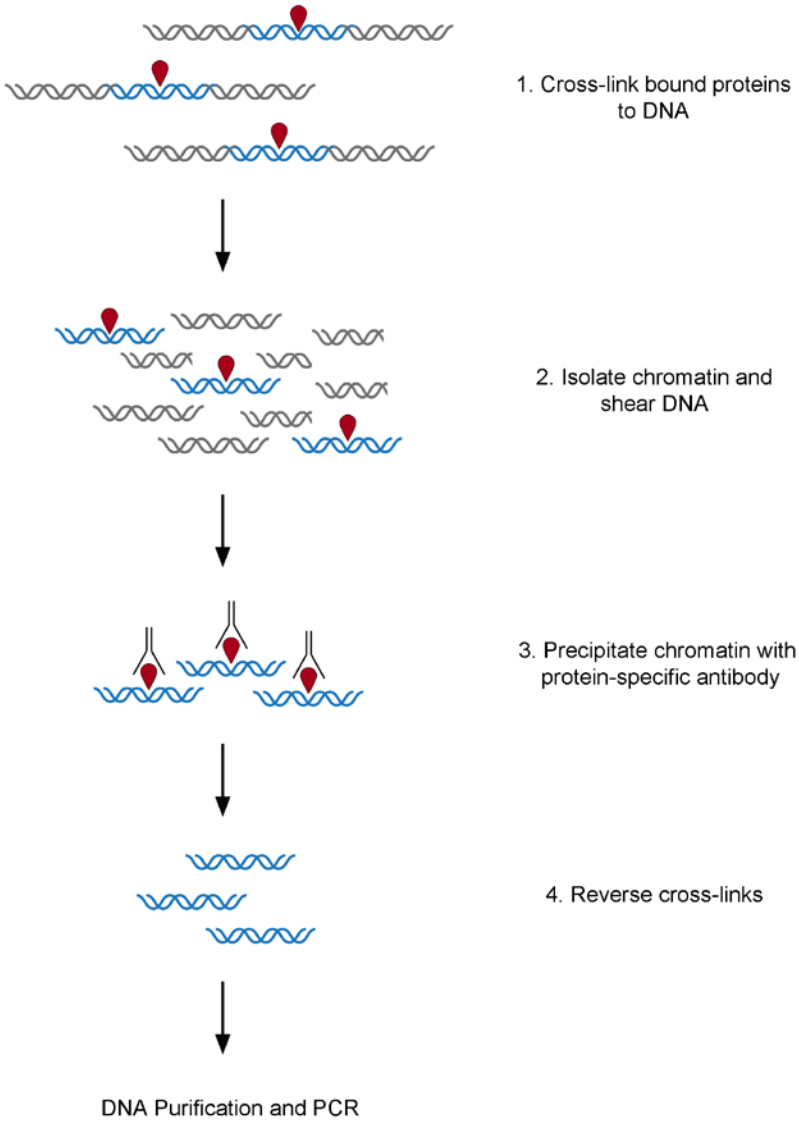


Figure 4-2: Overview of chromatin immunoprecipitation method

Protein:DNA complexes within cells are crosslinked using formaldehyde (final conc. 1%) then the samples are sonicated using a Soniprep to shear the chromatin to a target size of 200-800 bp. The protein of interest with its bound DNA is immunoprecipitated using protein G sepharose beads coupled to the antibody of interest then the cross-links are reversed and the enriched DNA is purified and prepared for analysis

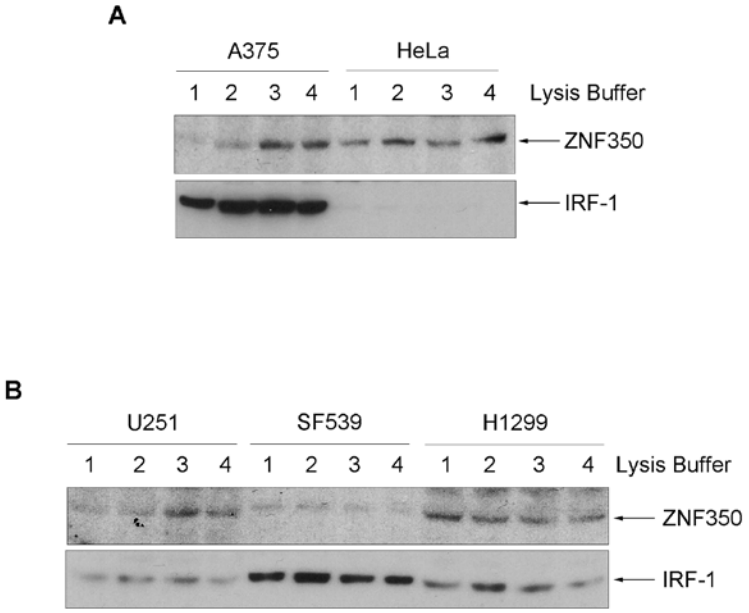


Figure 4-3: Optimization of cell type and nuclei lysis buffer

(A) Immunoblot of A375 and HeLa cells lysed in different ChIP lysis buffers (1-4, see 2.7.2 for recipes). Endogenous protein levels were analysed by 10% SDS-PAGE/immunoblot developed using anti-ZNF350 (abcam 77085) or anti-IRF-1 mAb (BD) antibodies. (B) As in (A) but with lysates from U251, SF539 and H1299 cells.

To determine with more accuracy the size of DNA fragments, I analysed my samples using the Agilent Bioanalyzer, performed by the Wellcome Trust Clinical Research Facility, which provides sizing, quantitation and quality control of DNA. To ensure efficient chromatin shearing, it is important that the proteins and DNA are not over-crosslinked and the ratio of cell number to lysate volume is at the most favourable level. In addition, the concentration of SDS in the lysis buffer can greatly affect the sonication efficiency and as such I found that 1% SDS is required. After optimising these steps, I then determined the number of cycles required to shear the chromatin to a size of 200-800 bp. When using the Bioruptor, the agarose gel indicates the fragment size is optimal after 60 cycles (30 seconds on, 30 seconds off) (Figure 4-4A, left panel), however the bioanalyzer trace reveals some quite large fragments in the sample (Figure 4-4A, right panel). Due to the length of time the samples need to be sonicated to obtain the correct DNA fragment size when using the bioruptor, the resulting kinetic energy released may make protein degradation more prevalent. The Soniprep tip sonicator, shears DNA in 10 second pulses with 30 second incubation on ice in between. From both agarose gel analysis and the bioanalyzer DNA size trace I ascertained that between 25-35 pulses of sonication resulted in the optimal fragment size (Figure 4-4B). More detailed analysis showed that 30 pulses of sonication resulted in the highest abundance of 200-800 bp fragments in relation to the whole sample, with an average size of 370 bp (Figure 4-5).

4.2.2.2 Immunoprecipitation

The quality of ChIP experiment is governed by the specificity of the antibody and the degree of enrichment achieved in the affinity precipitation step [286]. In order to select the most efficient antibody, I compared immunoprecipitation and enrichment specificity using immunoblotting and PCR with known gene targets of ZNF350 or IRF-1. Briefly, the protein-specific antibody was coupled to Protein G sepharose beads, as described in 2.6.3.2.1, and incubated with chromatin diluted in IP buffer, to reduce the concentration of SDS to 0.1%. The antibody:protein complex was washed extensively to remove non-specific binding and the specific protein-bound

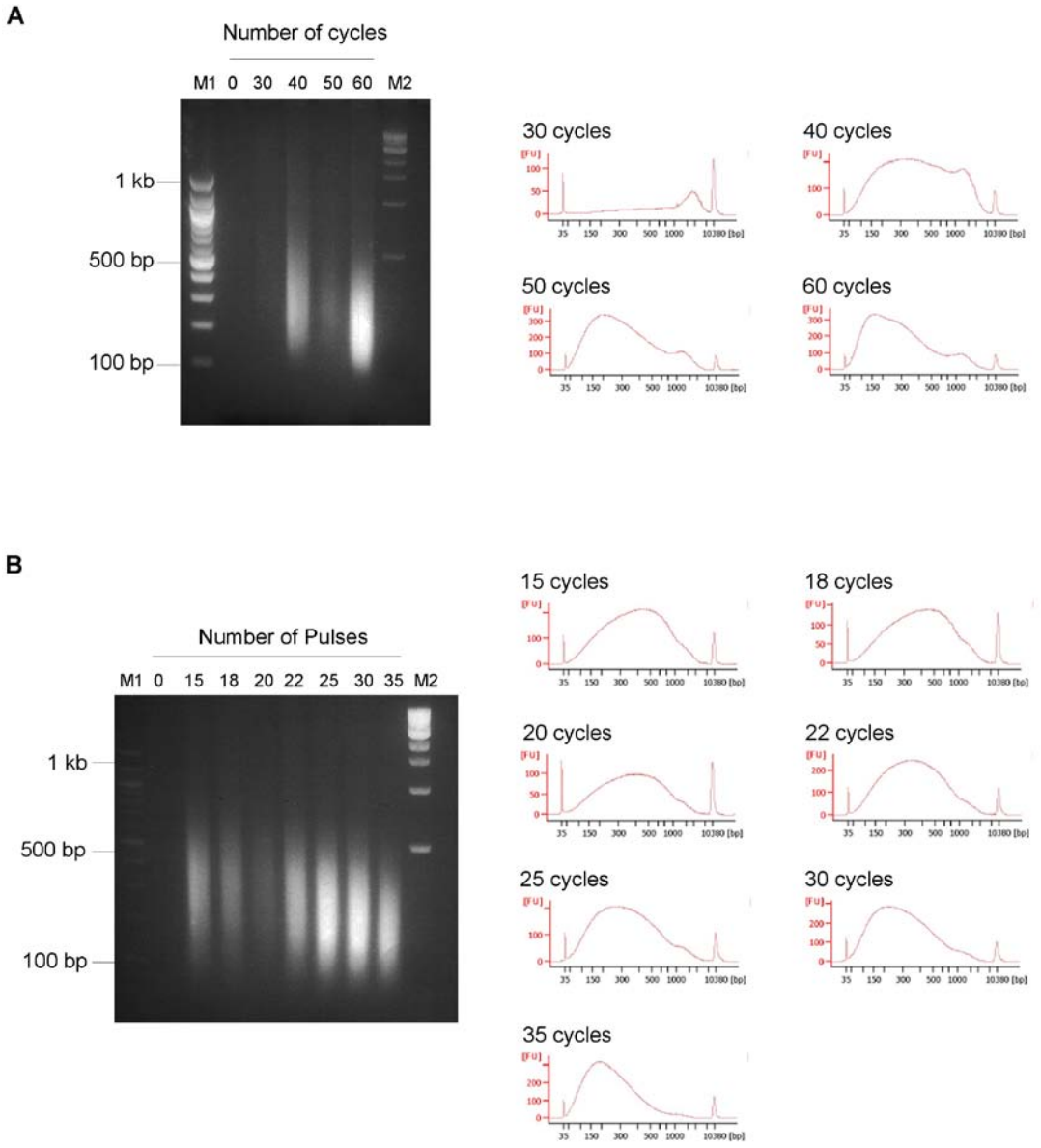
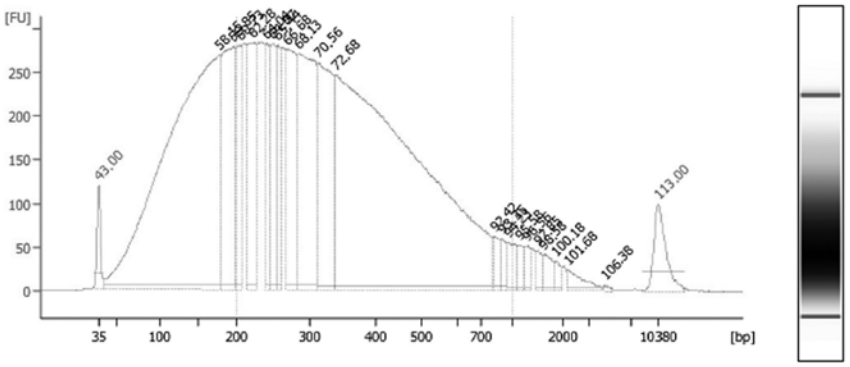


Figure 4-4: Optimization of chromatin shearing

(A) A375 cells were lysed then chromatin was sheared using a Bioruptor (UCD-200, Diagenode) for the specified number of cycles (30 sec ON, 30 sec OFF) and sizes measured by running on a 2% agarose gel (left panel) or analysed using a Bioanalyser (right panel; Agilent technologies; performed by the Wellcome Trust Clinical Research Facility, Edinburgh). (B) A375 cells were lysed then chromatin was sheared using a Soniprep 150 tip sonicator for the specified number of cycles (10 sec at amplitude 8-10 microns with 1 min incubations on ice in between). The sonicated samples were analysed as in (A; left panel: agarose gel, right panel: bioanalyser:)



Peak	Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	178	3,149.99	26,877.0	
3	195	741.24	5,745.4	
4	205	257.40	1,903.5	
5	222	440.65	3,010.3	
6	241	236.30	1,485.9	
7	250	248.42	1,505.8	
8	256	225.70	1,335.0	
9	270	427.87	2,403.8	
10	285	739.02	3,922.5	
11	313	549.35	2,655.6	
12	339	2,993.24	13,362.1	
13	828	36.48	66.7	
14	902	25.84	43.4	
15	971	43.51	67.9	
16	1,153	23.13	30.4	
17	1,307	20.32	23.6	
18	1,509	20.81	20.9	
19	1,623	29.18	27.2	
20	1,873	14.58	11.8	
21	2,210	28.32	19.4	
22	4,594	3.18	1.1	
23	10,380	75.00	10.9	Upper Marker

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]
200	1,000	9,027.2	63	370	39.6	6,650.13	32,236.8

Figure 4-5: distribution of DNA fragments using Soniprep 150 for 30 cycles

A375 cells were lysed and chromatin sheared using a Soniprep 150 tip sonicator for 30 cycles. DNA fragment size was analysed using a Bioanalyzer (Agilent technologies). Lower panel shows average size of DNA is 370 bp.

chromatin was then eluted. Following elution, the protein-DNA crosslinks were reversed and the DNA was purified then quantified.

At the time of this study, only one ChIP-grade anti-ZNF350 antibody (Abcam) existed, so I compared the efficiency of this antibody in extracting ZNF350-bound chromatin to that of a non-ChIP-grade antibody (Sigma). After the elution of ZNF350-bound chromatin a lot of protein remained bound to the antibody-bead complex so I performed two elution steps to ensure the most chromatin was extracted. To determine the abundance and specificity of extracted chromatin using the two antibodies, I generated primers flanking the reported ZNF350 binding site in the *ANG1* gene [52] and performed PCR. The resulting gene amplification was analysed on an agarose gel (Figure 4-6A). I found that immunoprecipitation using the ChIP-grade anti-ZNF350 antibody (Abcam) resulted in a much higher enrichment of the *ANG1* gene then the non-ChIP-grade antibody (Sigma). Similarly, I compared the immunoprecipitation efficiency of two anti-IRF-1 antibodies, H-205 (Santa Cruz) and C20 (BD) and analysed the enrichment of the known IRF-1 target, *TLR3* (Figure 4-6B). While the use of both IRF-1 antibodies results in the enrichment of the *TLR3* gene, the H-205 antibody appears to generate a higher abundance of *TLR3* in repeated experiments. Interestingly, the H-205 anti-IRF-1 antibody has also been used in a previous genome-wide analysis of IRF-1 binding in primary monocytes [273]. To confirm that the gene enrichment shown for *ANG1* and *TLR3* is ZNF350- and IRF-1 specific, respectively, I analysed the enrichment of *GAPDH,* a gene which has not been reported to bind either ZNF350 or IRF-1. Figure 4-6C shows that neither ZNF350- or IRF-1 extracted chromatin enriches for the *GAPDH* gene. In addition, by using the anti-IgG antibody to immunoprecipitate non-specific chromatin, the lack of gene enrichment for either *ANG1* or *TLR3* confirms the specificity of both anti-ZNF350 and anti-IRF-1 antibodies. Having successfully optimized ChIP, samples comprising of Input, ChIP-ZNF350 and ChIP-IRF-1, along with biological replicates for each condition, were sent off for library preparation and high-throughput sequencing using the HiSeq 2000 Illumina platform by the GenePool [287].

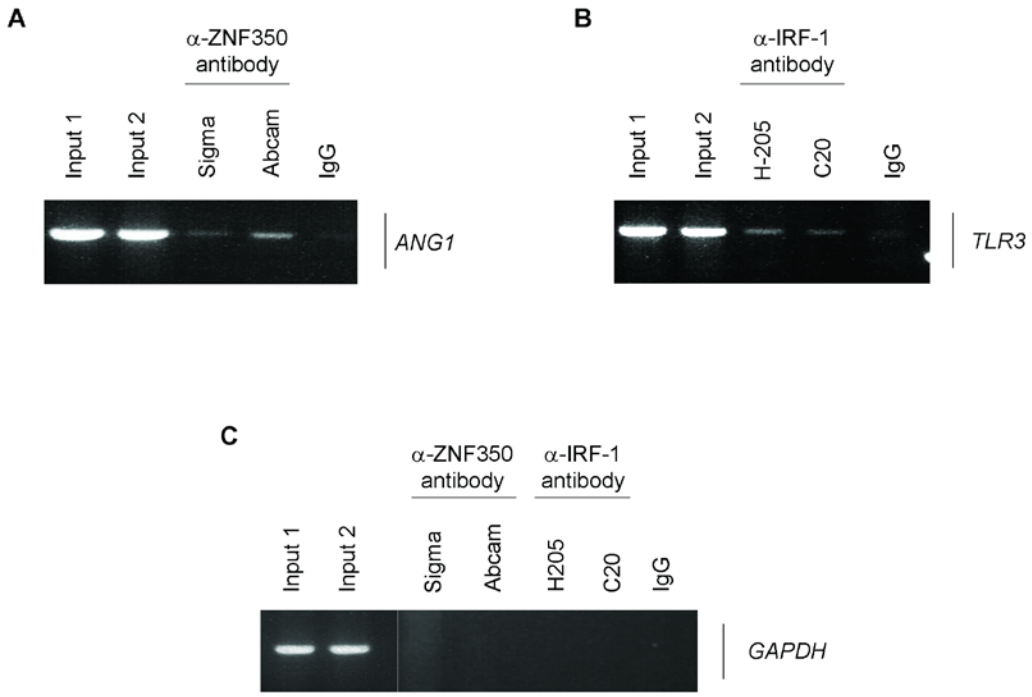


Figure 4-6: Verification of ChIP success

(A) ChIP performed using two different anti-ZNF350 antibodies, one produced by Sigma and one produced by Abcam (77085). Enrichment of DNA at a known ZNF350 target (*ANG1*) was analysed by PCR with primers designed around the ZNF350 binding site on the *ANG1* gene. Background enrichment determined by ChIP using anti-IgG antibody. (B) ChIP performed using H-205 (Santa Cruz) and C20 (BD) anti-IRF-1 antibodies and protein binding confirmed by PCR with primers designed around the IRF-1 binding site on the *TLR3* gene. Background enrichment determined by ChIP using anti-IgG antibody. (C) ChIP performed using antibodies described in (A) and (B) and non-specific binding analysed by PCR with primers designed on the *GAPDH* gene

4.2.3 High-throughput sequencing

High throughput sequencing, also known as next generation sequencing (NGS) was performed by the sequencing facility at The GenePool, University of Edinburgh which used the Illumina HiSeq platform to sequence 50 bp single end reads of the immunoprecipitated DNA. The Illumina HiSeq 2000 uses sequencing by synthesis (SBS) technology, an outline of the sample preparation and sequencing process is demonstrated in Figure 4-7. Briefly, unique adaptors are ligated to both ends of the fragmented DNA, of which 200-300 bp fragments are isolated by gel extraction and amplified using limited cycles of PCR. The flow cell surface is coated with single stranded oligonucleotides that correspond to the sequences of the adaptors ligated during the sample preparation stage. Single-stranded, adapter-ligated fragments are then bound to the surface of the flow cell and exposed to reagents for polymerase-based extension. Priming occurs as the free end of the ligated fragment "bridges" to a complimentary oligonucleotide on the surface of the flow cell. Repeated steps of denaturation and extension results in amplification of single molecules in millions of unique clusters across the flow cell surface. The flow cell is then loaded into the HiSeq sequencer for automated cycles of extension and imaging. The sequencing by synthesis technology adopted by Illumina uses a reversible terminator-based method that enables detection of single bases as they are incorporated into growing DNA strands. A fluorescently-labelled terminator is imaged as each dNTP is added, with a different fluorescent emission identifying which of the four bases is incorporated at that position. The terminator is then cleaved to allow incorporation of the next base with this process repeated until ultimately a series of images is generated, each representing a single base extension at a specific cluster [288-291].

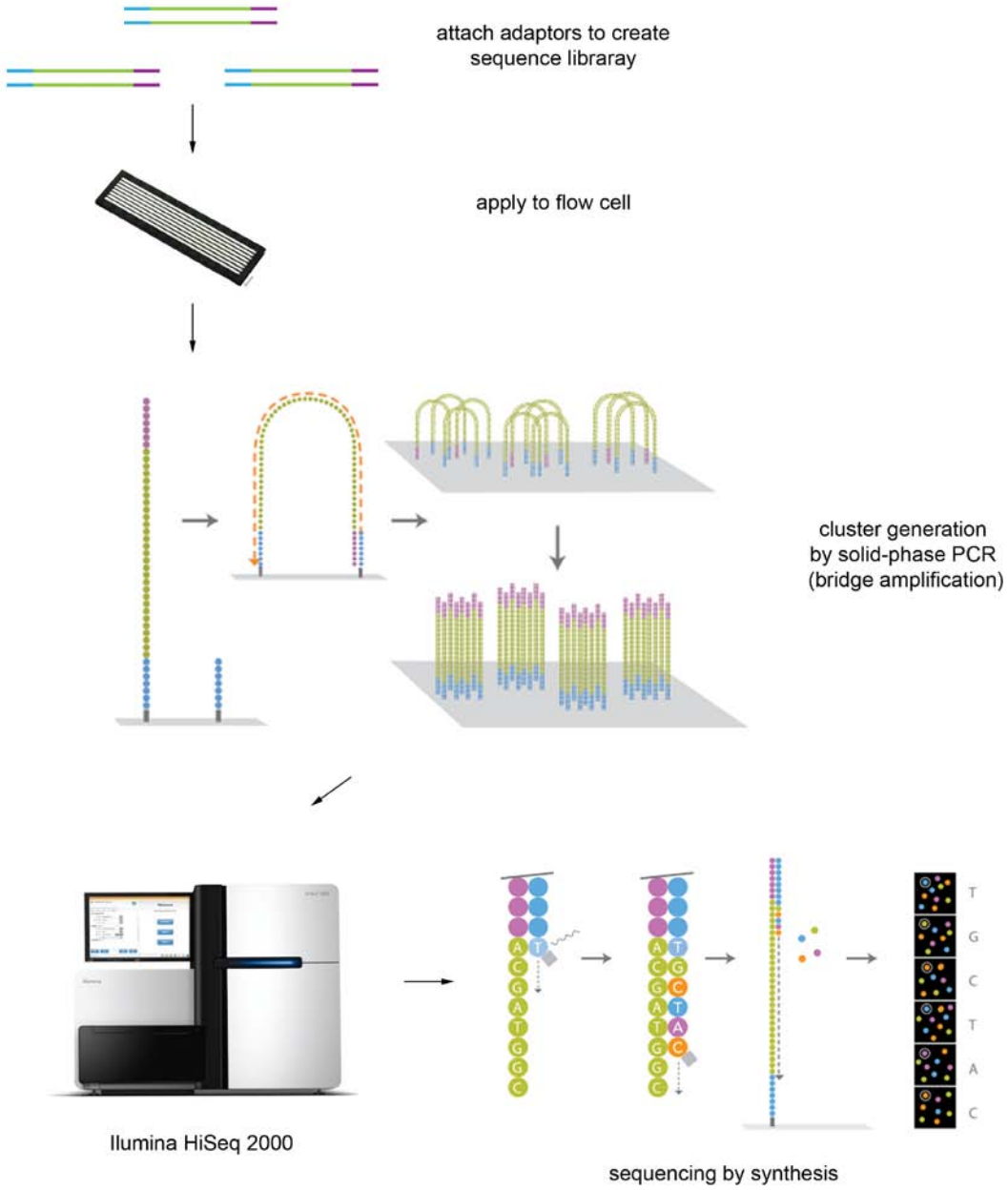


Figure 4-7: Illumina Sequencing by Synthesis technology

Methodology of high-throughput sequencing by synthesis using the Illumina Hiseq 2000 [292]. Sequencing performed by The Genepool, Edinburgh [156].

4.2.3.1 Analysis of ChIP-seq data

For the analysis of the raw sequencing data, I used the web-based tool Galaxy. Genome browsers, such as UCSC ([293], http://genome.ucsc.edu), NCBI [294] and Ensembl ([295], http://www.ensembl.org), allow experimental biologists with no programming experience to locate and visualize genomic regions using graphical interfaces; however, more sophisticated analyses still require programming and database skills. Galaxy is a system for the integration of genomic sequences, their alignments and functional annotation and was designed to allow users with no programming experience to carry out more advanced large scale genome analysis ([157], http://g2.bx.psu.edu). Galaxy is not a genome browser but instead allows users to gather and manipulate data in a variety of different ways then display the results using existing browsers, such as UCSC and Ensembl [157].

Galaxy has many tools for the analysis of ChIP-seq data and as a user with no programming experience, this enabled me to analyse my data. The goal of analysing ChIP-seq data is to perform successive steps to obtain a list of peaks demonstrating binding locations within the genome for the proteins of interest. A screenshot of the Galaxy web tool and the workflow used for the analysis of ChIP-seq data is demonstrated in Figure 4-8 (A and B, respectively). Briefly, the raw sequencing read files for two Input replicates, two ZNF350 replicates and two IRF-1 replicates were uploaded to Galaxy and basic quality statistics were computed for each file. These quality scores were used to generate box plots for the Input (Figure 4-9), ZNF350 (Figure 4-10) and IRF-1 (Figure 4-11) replicates, which show that for all six samples the sequencing quality is very good due to the quality score for most bases is 40 on the quality score index. Mapping of the reads was then performed on the reference genome, in this case human hg19 genome, with the BWA tool [296] to obtain their coordinates. Table 4-1 shows the total number of reads for each sample and the percentage of unmapped reads to the reference genome. Finally peak-calling using MACS (Model-based analysis of ChIP-seq) looks for fold enrichment of the sample over input, or expected background. MACS also uses a dynamic Poisson distribution to effectively capture local biases in the genome, allowing for more robust predictions [297].

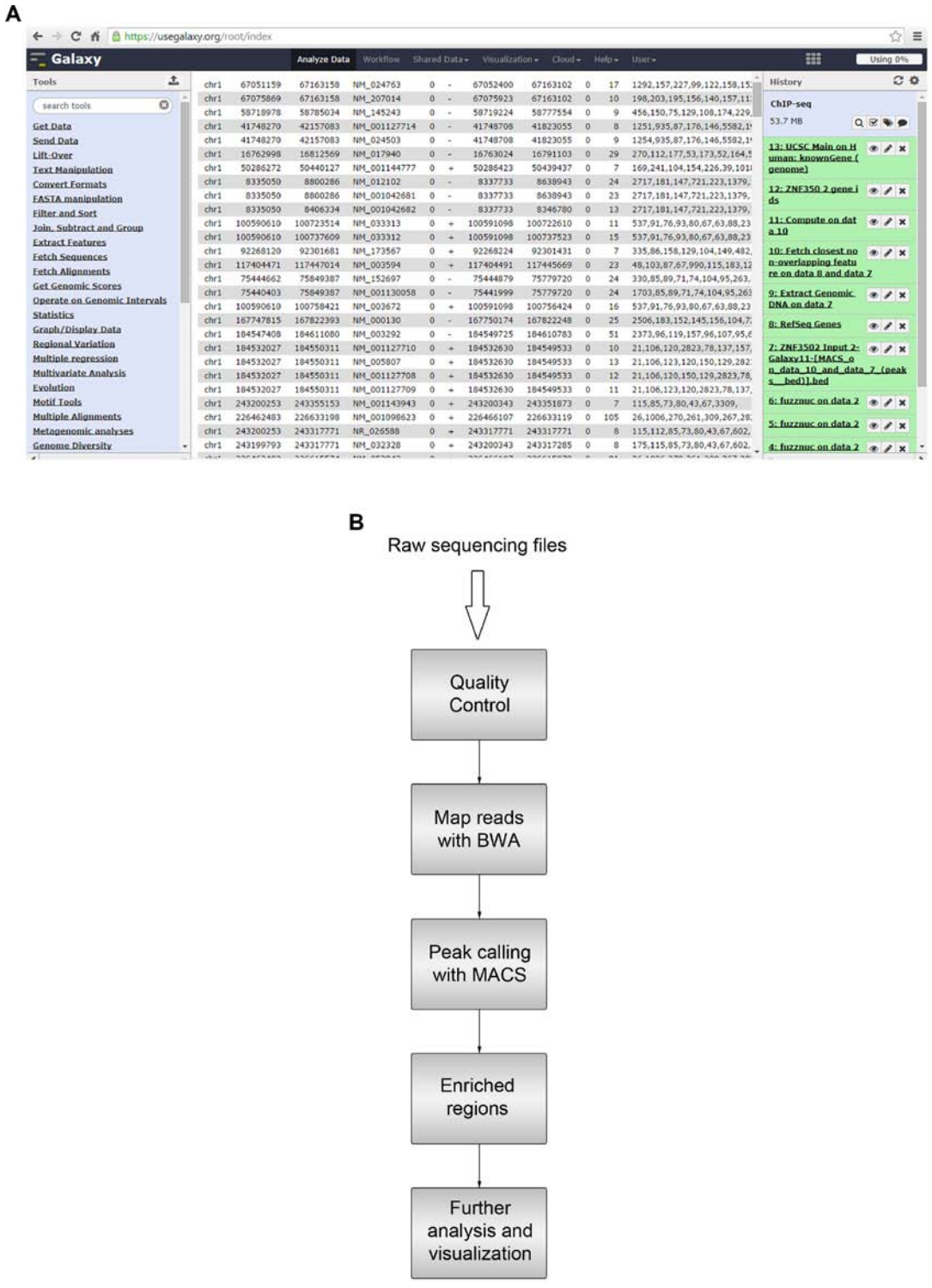


Figure 4-8: ChIP-seq analysis using the Galaxy platform

(A) Screenshot of Galaxy web tool. The tools are located in the left panel, and the workflow history is located in the right panel. The data can be viewed in the central panel. (B) ChIP-seq analysis workflow.

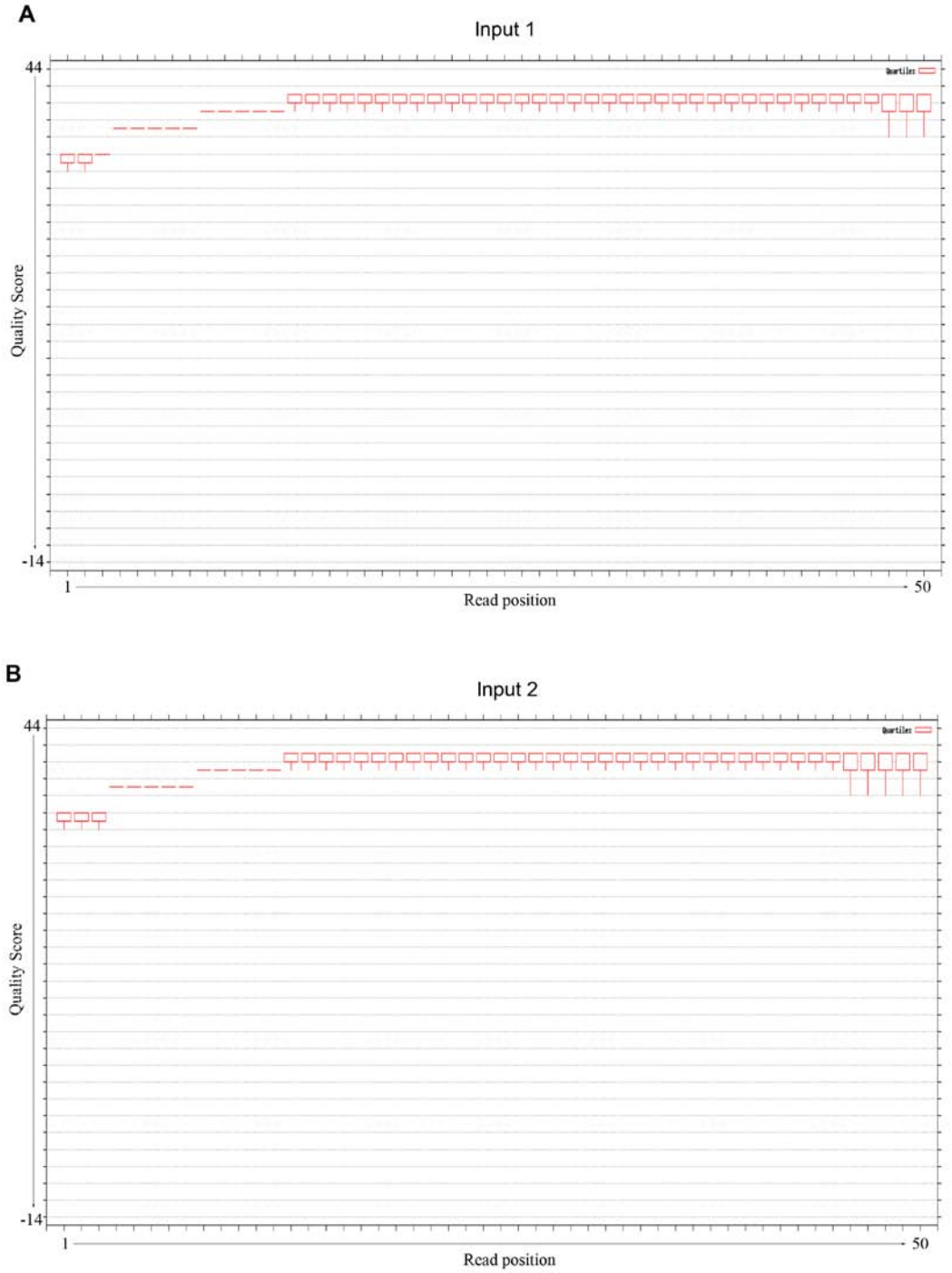


Figure 4-9: Input replicates sequence quality analysis

(A) Quality score boxplot for Input 1 replicate generated from compute quality statistics tool in Galaxy. A quality score above 30 represents 99.9% accuracy of base position whereas a score above 40 represents 99.99% accuracy. (B) As in (A), except quality statistics for Input replicate 2.

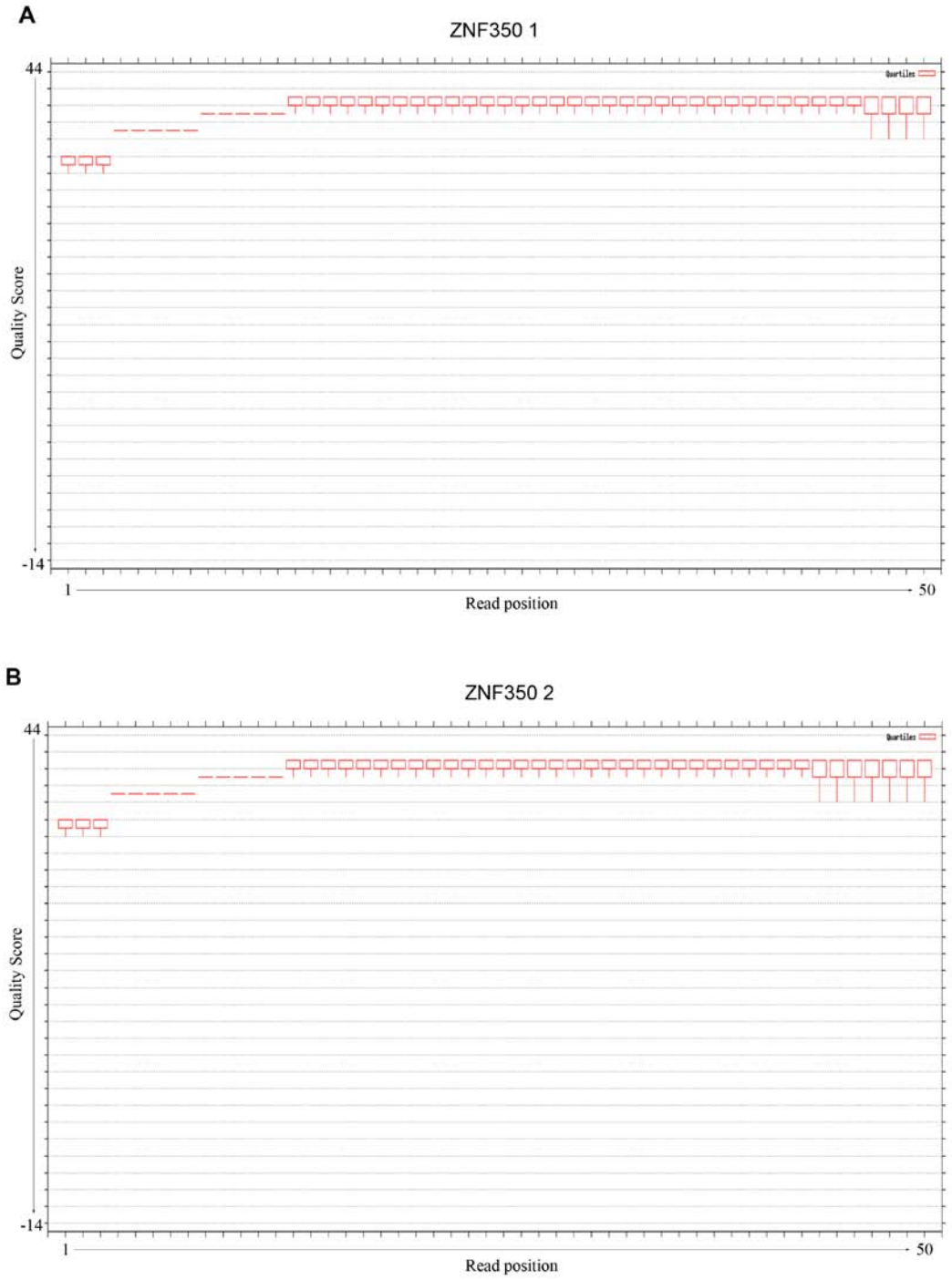


Figure 4-10: ZNF350 replicates sequence quality analysis

(A) Quality score boxplot for ZNF350 replicate 1 generated from compute quality statistics tool in Galaxy. A quality score above 30 represents 99.9% accuracy of base position whereas a score above 40 represents 99.99% accuracy. (B) As in (A), except quality statistics for ZNF350 replicate 2.

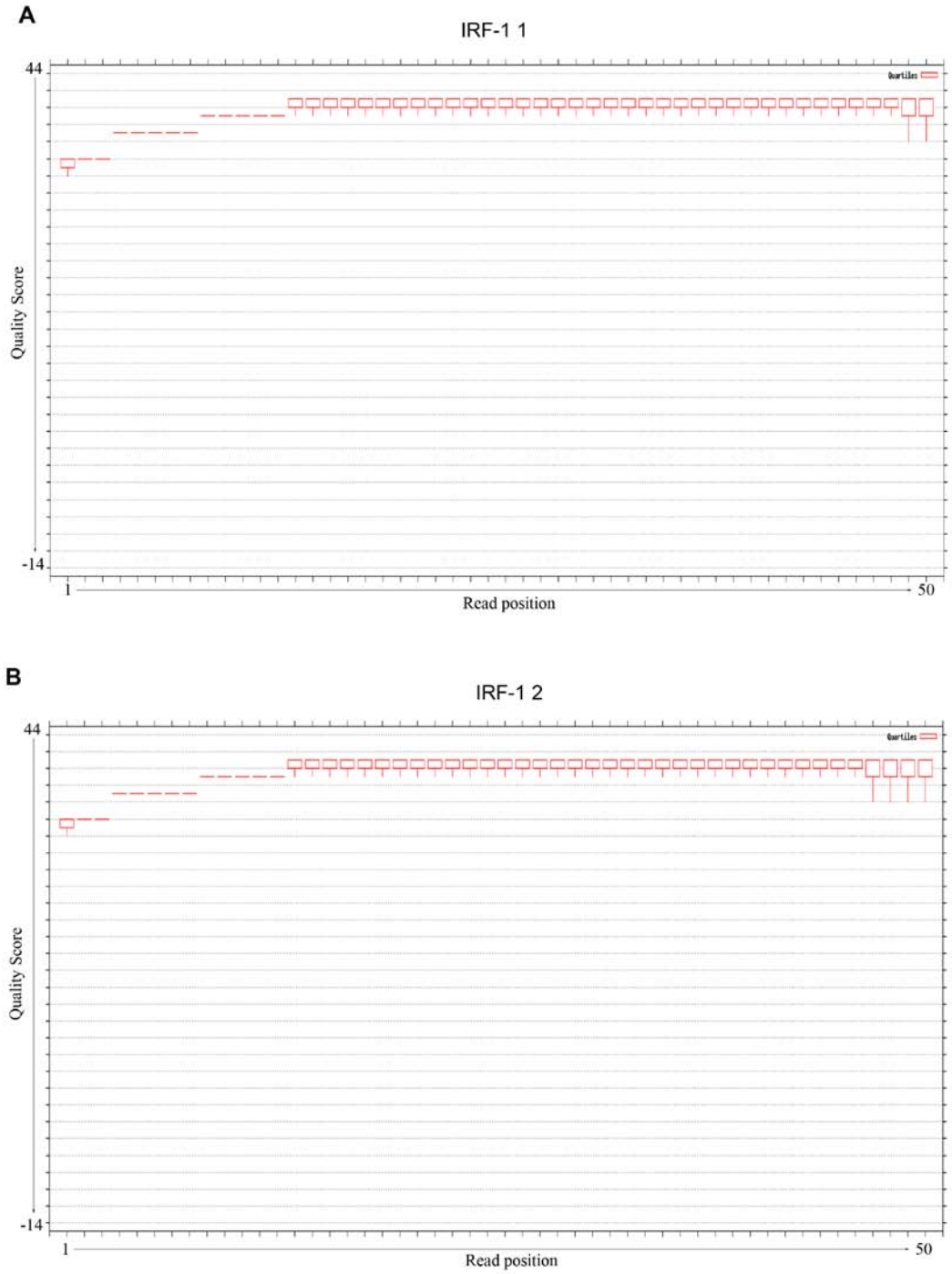


Figure 4-11: IRF-1 replicates sequence quality analysis

(A) Quality score boxplot for IRF-1 replicate 1 generated from compute quality statistics tool in Galaxy. A quality score above 30 represents 99.9% accuracy of base position whereas a score above 40 represents 99.99% accuracy. (B) As in (A), except quality statistics for IRF-1 replicate 2.

Table 4-1: Mapping of reads by BWA

	Total number of reads	***Percentage of unmapped reads***
Input 1	44,028,789	4%
Input 2	29,625,888	7%
ZNF350 1	24,143,592	4%
ZNF350 2	32,448,896	9%
IRF-1 1	115,909,187	4%
IRF-1 2	28,606,926	7%

4.2.3.1.1 Comparison of biological replicates

The data analysis tools on Galaxy generate a file for each biological replicate of the protein of interest containing peaks of enriched regions of the genome to which the protein of interest binds. The files generated for the Input control replicates show peaks representing background binding (Table 4-2).

Table 4-2: Number of peaks called for each replicate

	Number of called peaks
Input 1	1982
Input 2	1763
ZNF350 1	1088
ZNF350 2	478
IRF-1 1	258
IRF-1 2	989

It can be seen that there is quite a difference between the number of peaks generated for each replicate with ChIP-ZNF350 1 generating over twice as many peaks compared to ChIP-ZNF350 2 while ChIP-IRF-1 2 gives over three times as many peaks as opposed to ChIP-IRF-1 1. Despite this discrepancy, I proceeded to use

Galaxy tools to intersect the two replicates for each condition to identify any overlapping peak locations. While the Input replicates showed over 50% overlap (Figure 4-12A), unfortunately the number of common peaks found between ZNF350 replicates is 10, whilst only 9 peaks overlap between the two IRF-1 replicates (Figure 4-12B and C, respectively). Usually successful ChIP-seq replicates show considerable overlap which gives significant weight to the results obtained from ChIP-seq; however the results presented here would indicate that perhaps a problem has occurred. Due to the sequence analysis showing high quality for all samples (Figures 4-9, 4-10 and 4-11), the lack of overlapping peaks between replicates are most likely to be attributed to problems during the ChIP experimental process. Moreover, since the input replicates which were not subject to immunoprecipitation during chromatin isolation show considerable overlapping peaks compared to ZNF350 and IRF-1 samples, this would suggest there may have been a problem with the immunoprecipitation of chromatin using the anti-ZNF350 and anti-IRF-1 antibodies.

To try to further identify the discrepancies between replicates, a bioinformatition with the Semple Group Philippe Gautier, performed a further ChIP-seq quality check using HOMER (Hypergeometric Optimization of Motif EnRichment) [298]. HOMER is a compilation of command line programs for unix-style operating systems and contains various tools for analyzing ChIP-seq data, such as tag autocorrelations [298]. As described previously, ChIP works by shearing chromatin to isolate fragments of DNA bound by a specific protein. Due to this, and the requirements for high-throughput sequencing, only DNA fragments of a certain size are sequenced. The size of these fragments plays an important role in extracting significant data and being able to accurately determine the location of protein binding sites. HOMER [298] performs an autocorrelation analysis of tag positions to estimate the length of the fragments used for sequencing. By calculating the distance from one read to another along the same chromosome, a histogram of these distances can be plotted [298]. This type of quality analysis is independent of peak calling and

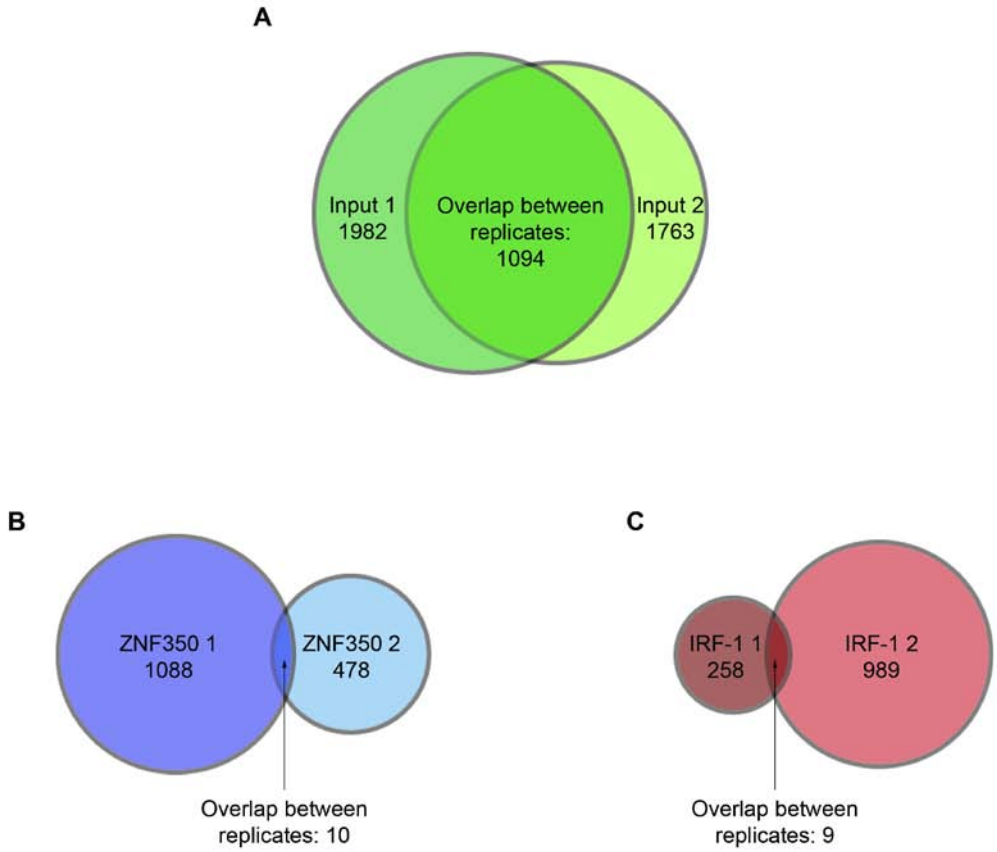


Figure 4-12: Peaks called for ZNF350 and IRF-1 replicates

(A) Venn diagram depicting the number of peaks called for each Input biological replicate and the peaks which are found in both replicates. (B) As in (A) but overlapping peaks of ZNF350 replicates (C) As in (A) but overlapping peaks of IRF-1 replicates.

is based on the observation that in a good quality ChIP-seq experiment, sites bound by the protein of interest show a significant cluster of enriched DNA sequence tags centered around the binding site on the forward and reverse strands [286]. The location of these sequence tags in relation to the binding site center depends on the distribution of the fragment size [299]. This shifting of accumulated tags between strands is not observed in control experiments, such as input DNA, which has made it possible to quantify IP enrichment based on the correlation between genome-wide stranded tag densities [286]. Typically, two peaks are generated in high-quality ChIP-seq data sets: a larger peak of enrichment corresponding to the fragment length and a smaller peak corresponding to read length [298]. The principle of an autocorrelation plot is demonstrated in Figure 4-13A and an example of an autocorrelation plot representing a successful ChIP-seq experiment shown in Figure 4-13B.

Autocorrelation plots generated by Philippe Gautier for ZNF350 and IRF-1 ChIP replicates (Figure 4-13C and D, respectively) show that the peak representing the read length (in this case 50 bp) is much larger than the peak corresponding to the fragment length. This second peak is very small for both ZNF350 replicates (Figure 4-13C) and IRF-1 2 (Figure 4-13D, right panel) while IRF-1 1 is the only sample that perhaps comes close to the 'model' plot (Figure 4-13D, left panel). This data along with the lack of overlapping peaks between replicates would indicate that there may have been problems with the ChIP part of this study. The experimental design for ChIP involved the pooling of chromatin from multiple cell lysates, which was then divided equally between the different protein-specific IP's. This resulted in identical biological material being sampled, which would therefore indicate that the efficiency of the IP could be responsible for differences observed in protein binding. Despite extensive optimization, the ChIP-grade antibodies used in this study were both polyclonal which, although are usually thought to work well in ChIP because they contain a mix of antibodies that recognize several different epitopes on the same protein to achieve a greater amount of sensitivity (i.e. pull-down power), this sometimes sacrifices specificity in the process [300]. Additionally, genome-wide experiments are intrinsically subject to noise, and thus replicate experiments

systematically produce different values or ranks of peaks, even if the samples are of high quality. For this reason, if two replicates are independently thresholded at an identical value, peaks that are above the threshold in one might be below the threshold in the other, and vice versa. Although this is necessary to confidently define a stringent set of peaks, it excludes a reasonable estimation of perhaps borderline significant peaks that are shared between replicates [301]. However, the fundamental issue with the results presented here is not the total number of peaks but the very small overlap between ZNF350 replicates and IRF-1 replicates. Indeed, when different algorithms were used to call peaks (MACS 1.4, MACS 2.0 and SISSRs, performed by Philippe Gautier) there was some variability between the different programs. However, the effect of these disparities on the final result, i.e. overlap between replicates, was negligible (data not shown).

4.2.3.1.1.1 ZNF350

Despite the ChIP-seq data being of unsatisfactory quality to gain any significant insight into the binding properties of ZNF350 and IRF-1, I decided to compare the biological replicates for each sample to identify if they share common aspects of binding locations. Table 4-3 lists the 10 ZNF350 peaks shared between the biological replicates, detailing the corresponding gene to each potential binding region. Interestingly, three shared peaks are located in *ZNF350* itself whilst another three are found in *MDM2*, one in *TP53* and one in *IRF-1*.

Table 4-3: Overlapping peaks for ZNF350 replicates

Peak number	***Chromosome***	***Start***	***End***	***Associated gene***
222	chr12	69222558	69222709	*MDM2*
223	chr12	69229617	69229765	*MDM2*
224	chr12	69233294	69233440	*MDM2*
310	chr17	7572813	7572970	*TP53*
331	chr17	80254149	80254267	-
366	chr19	52468071	52469469	*ZNF350*
367	chr19	52471827	52471916	*ZNF350*
368	chr19	52472269	52472430	*ZNF350*
654	chr5	131822643	131822824	*IRF-1*
791	chr9	10234	10345	-

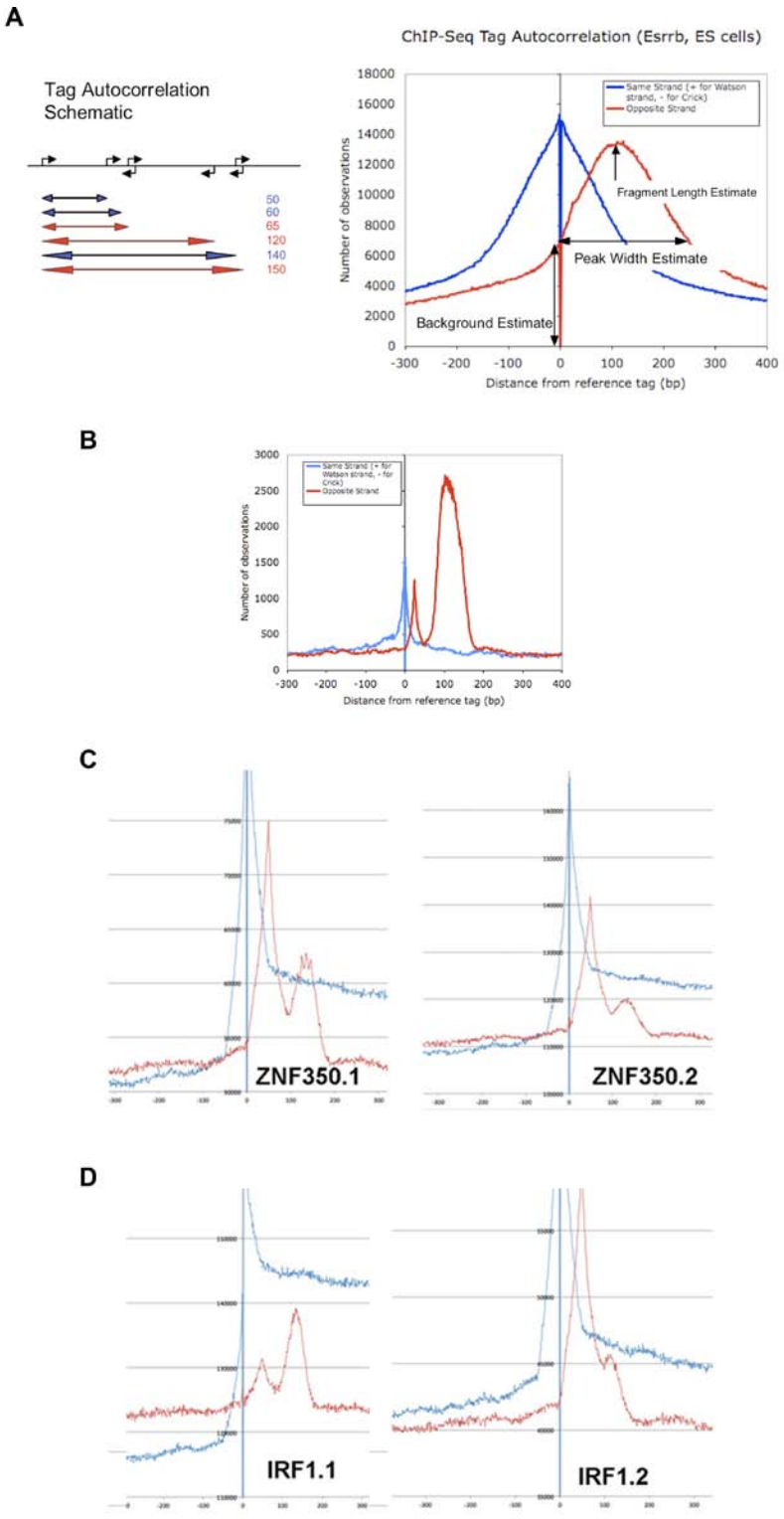


Figure 4-13: Autocorrelation plots of ChIP-seq data sets

(A) Principle of an autocorrelation plot. (B) Example of an 'ideal' plot, both from HOMER [298]. (C) ZNF350 replicates autocorrelation plots. (D) IRF-1 replicates autocorrelation plots.

In order to identify general differences in the peaks called for the two ZNF350 replicates, I sought to establish the location of the peaks throughout the genome. Genomic Regions Enrichment of Annotations Tool (GREAT) associates genomic regions with genes by defining a 'regulatory domain' for each gene in the genome [302]. Each genomic region is associated with all genes in the regulatory domain it lies within. The regulatory domain assigned by GREAT consists of a basal domain that extends 5 kb upstream and 1 kb downstream from its transcription start site and is extended up to the basal regulatory domain of the nearest genes within 1 Mb upstream or downstream. In addition, GREAT further defines the regulatory domains of a handful of genes by incorporating their experimentally determined regulatory domains [302]. GREAT can also characterise the genes associated with the genomic regions in terms of cellular pathways. Pathway commons ontology contains data on pathways from multiple sources including biochemical reaction, complex assembly, transport and catalysis events as well as physical interactions involving proteins, DNA, RNA, small molecules and complexes [302].

By inserting the ZNF350 peak coordinates into the GREAT tool, several graphs are generated which demonstrate their location within the genome. The "number of associated genes per region" graph shows how many genes each genomic region is assigned as possibly regulating based on the rule described above [302]. For ZNF350 replicate 1, 22.5% of the peaks are not associated with any gene whereas only 5.4% of peaks called for ZNF350 replicate 2 have no associated gene regions (Figure 4-14A). Of the ZNF350 1 peaks assigned to gene regions, 45.8% are associated with 1 gene whereas 31.7% are associated with 2 genes. This pattern is not found for the second ZNF350 replicate where only 17.8% of peaks are assigned to 1 gene with the majority (76.8%) being associated with 2 genes (Figure 4-14A). GREAT also produces graphs showing the distance between the peak locations and the transcription start site (TSS) of the putatively regulated genes. The majority of both ZNF350 replicate peaks are located 50-500 kb upstream or downstream from the TSS of the associated gene with very few peaks located within the promoter regions of genes (0-5 kb upstream of the TSS; Figure 4-14B).

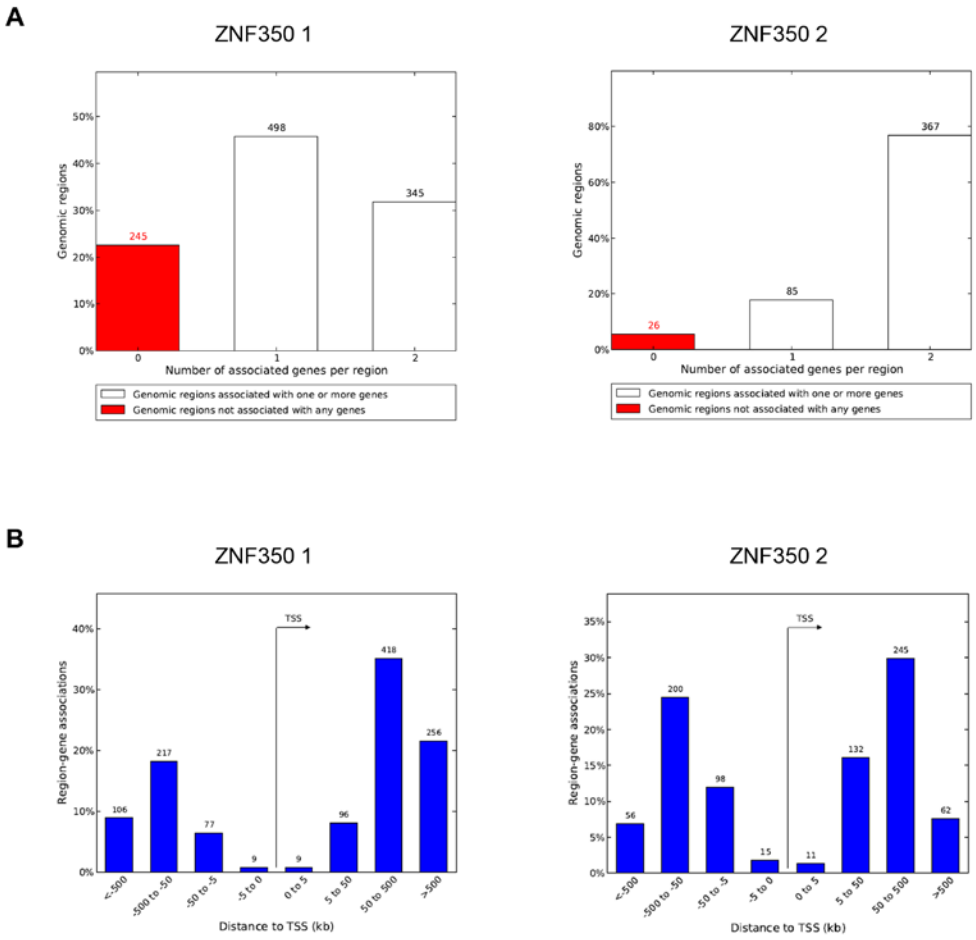


Figure 4-14: Comparison of binding regions of ZNF350 replicates

(A) GREAT graph showing the number of associated genes per genomic region to which a ZNF350 peak is observed. (B) GREAT graph showing the distance from the ZNF350 peak region to the transcription start site (TSS) of the associated gene. Since some ZNF350 peaks are associated with more than one gene, GREAT measure the distance to the TSS of all genes which accounts for the greater number of region-genes associations than in (A).

No significant over-representation of ZNF350 peaks were found in any of the specific pathways characterized by GREAT. Despite the binding pattern of the two ZNF350 replicates being similar in relation to the distance from the TSS, the number of genes per genomic region each replicate is associated with is quite different. This would demonstrate that the features of ZNF350 binding to the chromatin differs between replicates, indicating a problem with the ChIP experiment and unreliability of the data.

For each ZNF350 replicate, I sought to examine the specific sites bound by the protein by analysing the peak sequences for common motifs. Since the ZNF350 consensus motif (GGGxxxCAGxxxTTT) is not in any transcription factor binding site (TFBS) databases, I used the 'peak-motifs' tool on the Regulatory Sequence Analysis Tools (RSAT) [303; 304] website to obtain a list of overrepresented oligonucleotides. Unfortunately, there were no common motifs found to be shared between the replicates and none corresponing to the putative ZNF350 consensus binding sequence. Several consensus motifs for other transcription factor binding sites were identified in ZNF350 1 (Table 4-4) and ZNF350 2 (Table 4-5) peaks but again, none were significant or were shared in common with both replicates.

Table 4-4: ZNF350 1 identified TFBS motifs

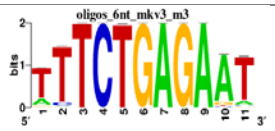
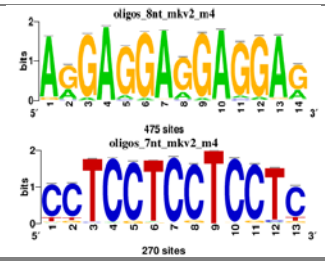
Motif	***Number of peaks with at least one predicted site***	***TFBS in JASPAR database***
	318 (29.23%)	Stat5a::Stat5b
	262 (24.08%)	SPI1, ETS1, ELF5

Table 4-5: ZNF350 2 identified TFBS motifs

Motif	*Number of peaks with at least one predicted site*	*TFBS in JASPAR database*
	145 (30.33%) 92 (19.25%)	ZNF263 SP1

4.2.3.1.1.2IRF-1

As mentioned previously, the number of peaks called for each IRF-1 replicate differs quite a bit in that ChIP-IRF-1 1 produced only 258 peaks whereas ChIP-IRF-1 2 resulted in 989 peaks (Table 4-1). When looking at the autocorrelation of peaks, IRF-1 replicate 1 showed a slightly better plot compared to IRF-1 2 (Figure 4-13D), however neither demonstrated an ideal autocorrelation plot representative of a successful ChIP-seq experiment. Despite the obvious troubles with the sequencing data, I sought to identify any overlapping genomic regions bound by both IRF-1 ChIP replicates. Table 4-6 lists the 9 IRF-1 peaks shared between the biological replicates, detailing the corresponding gene to each potential binding region. Interestingly, the two IRF-1 replicates both bind to regions associated with *MDM2, ZNF350,* and *IRF-1* itself, all genes which were found to be shared in common between the two ZNF350 replicates.

Table 4-6: Overlapping peaks for IRF-1 replicates

Peak number	*Chromosome*	*Start*	*End*	*Associated gene*
16	chr1	121394481	121395235	-
19	chr1	121450624	121451455	-
62	chr12	69222543	69222724	*MDM2*
63	chr12	69229579	69229776	*MDM2*
64	chr12	69233087	69233623	*MDM2*

115	chr19	52468021	52469485	*ZNF350*
116	chr19	52471747	52471997	*ZNF350*
157	chr4	186988522	186988688	-
169	chr5	131822633	131822841	*IRF-1*

As with the ZNF350 replicates, I sought to compare IRF-1 samples to identify if they demonstrated a similar binding pattern within the genome. Using GREAT, I compared the number of associated genes per region to which IRF-1 peaks are located within and found that both replicates demonstrate a similar pattern of association (Figure 4-15A). 8.1% of IRF-1 1 peaks and 14.4% of IRF-1 2 peaks are not associated with any genes whilst the majority of peaks for both replicates are found to be associated with two genes (61.6 % for IRF-1 1 and 52.2% for IRF-1 2; Figure 4-15A). Both IRF-1 replicates display a similar pattern in relation to the distance of the peaks from the TSS of the related gene in that the majority are found 50-500 kb upstream or downstream (Figure 4-15B). Interestingly the least frequent peak to TSS distance was 0-5 kb, the region representing the promoter. This pattern of binding is also comparable to that displayed by both ZNF350 replicates (Figure 4-14).

To explore the IRF-1 binding regions in greater detail, I sought to determine if the peak regions contained the IRF-1 consensus motif (Table 4-7) or any other TFBS in common using the RSAT 'peak-motifs' tool [303; 304]. For IRF-1 1, there were no significant overrepresented motifs and none corresponded directly to the JASPAR IRF-1 site. However, sites representing other TFBS that show some similarity to the IRF-1 motif are found in a small proportion of peaks (Table 4-8). Similarly, IRF-1 2 peaks do not contain any significant motifs and none with the known IRF-1 site although a few other TFBS were identified such as SPI1 and Sox3, which vaguely resemble the IRF-1 site (Table 4-9). Notably, an identified Stat5a:Stat5b TFBS in approximately 20% of IRF-1 2 peaks was also found in about 30% of ZNF350 1 peaks (Table 4-9 and Table 4-4, respectively). Since the known IRF-1 motif is not found to be overrepresented in the IRF-1 replicates, this is consistent with the idea that the ChIP-seq data is not of suitable quality to gain any novel insights.

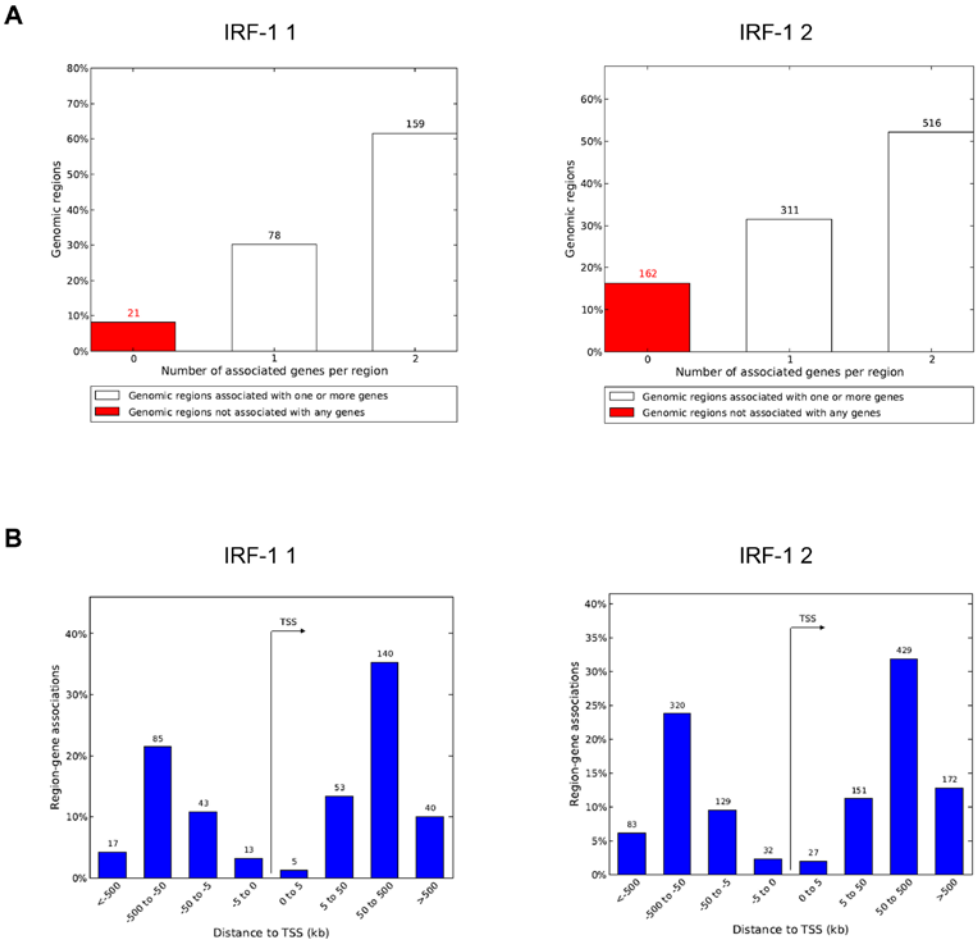


Figure 4-15: Comparison of binding regions of IRF-1 replicates

(A) GREAT graph showing the number of associated genes per genomic region to which a IRF-1 peak is observed. (B) GREAT graph showing the distance from the IRF-1 peak region to the transcription start site (TSS) of the associated gene. Since some IRF-1 peaks are associated with more than one gene, GREAT measure the distance to the TSS of all genes which accounts for the greater number of region-genes associations than in (A).

Table 4-7: JASPAR IRF-1 consensus motif

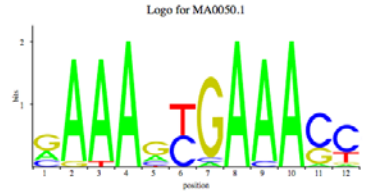
Motif	*TFBS*	*JASPAR ID*
	IRF-1	MA0050.1

Table 4-8: IRF-1 1 identified TFBS motifs

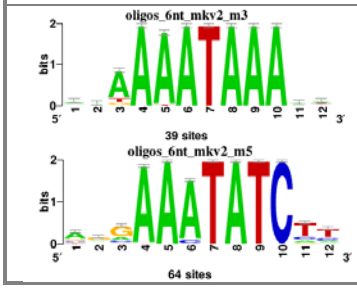
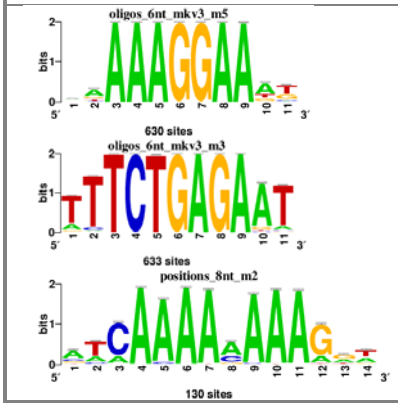
Motif	*Number of peaks with at least one predicted site*	*TFBS in JASPAR database*
	34 (13.18%) 29 (11.24%)	MEF2C, FOXL1, FOXI1 FEV, SPI1, GATA3

Table 4-9: IRF-1 2 identified TFBS motifs

Motif	*Number of peaks with at least one predicted site*	*TFBS in JASPAR database*
	262 (26.49%) 204 (20.63%) 170 (17.19%)	SPI1, ELF5, FEV Stat5a::Stat5b Sox3, Foxd3, Sox2

4.2.3.1.1.3Common peaks

After looking at the ZNF350 and IRF-1 peaks individually and while they show dissimilarity in the number of associated gene regions and consensus sequences present, there appeared to be some similarities in the location of the peaks within the genome. Indeed, by comparing the 10 overlapping ZNF350 replicate sites (Table 4-3) with the 9 overlapping IRF-1 replicate sites (Table 4-6), it can be seen from Table 4-10 that 6 regions are shared in common, corresponding to 3 peaks in *MDM2*, 2 peaks in *ZNF350* and 1 peak in *IRF-1*.

Table 4-10: Overlap between ZNF350 and IRF-1 peaks

ZNF350 peak number	***IRF-1 peak number***	***Chromo***	***Start***	***End***	***Associated gene***
222	62	chr12	69222543	69222724	*MDM2*
223	63	chr12	69229579	69229776	*MDM2*
224	64	chr12	69233087	69233623	*MDM2*
366	115	chr19	52468021	52469485	*ZNF350*
367	116	chr19	52471747	52471997	*ZNF350*
654	169	chr5	131822633	131822841	*IRF-1*

Despite the lack of confidence in the data, it is quite interesting to note that ZNF350 and IRF-1 peaks are found in the coding regions of both *ZNF350* and *IRF-1* genes. Further studies however, would be required to determine if this binding does actually occur in vivo and is not an artefact of the ChIP-seq data. Another notable finding is the location of ZNF350 peaks in the p53 (*TP53*) gene and both ZNF350 and IRF-1 peaks in the *MDM2* (*murine double minute 2*) gene. p53 is a transcription factor that has been show to play a major role in tumour suppression and the cell's response to oncogenic stresses [305-307]. It is one of the most commonly mutated proteins in human cancers, with around 50% of all malignancies harbouring a mutation in the p53 gene. p53 is known to be essential for preventing cancer development as it has been found to control over 100 target proteins involved in regulation of cell cycle, growth control, apoptosis, DNA repair, metastasis, angiogenesis and protein degradation in response to oncogenic and other cellular stress signals [308-310]. p53 activity is tightly controlled by the rate of its degradation through the ubiquitin-

proteasome pathway, with MDM2 being the main E3 ligase involved in this. Under normal conditions, p53 levels are controlled by complex system of feedback loops with dysregulation leading to the excessive ubiquitination and degradation of the p53 protein. This prevents its function as a tumour suppressor and results in uncontrolled cell growth and cancer development ([311] and references in [312]). It has recently been shown by the Ball group that the DNA binding domain of IRF-1 is ubiquitinated by MDM2, which acts by binding to the Mf2 domain. Furthermore, IRF-1 is only available for MDM2 binding and subsequent ubiquitination when in its DNA-unbound state, suggesting that IRF-1 degradation may be partly controlled by its ability to bind DNA [313]. IRF-1 and p53 are known to act together to induce p21 expression in response to DNA damage [114], however the exact relationship between the two proteins along with MDM2 has yet to be fully understood.

ZNF350 is known to repress the activity of the *GADD45A* gene by binding to intron 3 in conjunction with BRCA1 [43]. Following genotoxic stress, p53 binds directly to a consensus sequence located in intron 3 of the *GADD45A* gene and indirectly, through its interaction with other transcription factors, to the promoter region resulting in the activation of *GADD45A* transcription [314; 315]. After undertaking further studies to establish if ZNF350 does indeed bind to the p53 gene *in vivo,* it would also be of interest to examine if there is a relationship between ZNF350 and p53 regulation of *GADD45A* gene activity.

Although the sequencing data from the ChIP experiments does not appear to be particularly reliable due to the differences between the replicates, the possibility that ZNF350 and IRF-1 proteins bind to sites in their respective genes as well as *p53* and/or *MDM2* is quite intriguing. Further experiments will be required to validate *MDM2* and *TP53* as binding targets of ZNF350 and/or IRF-1 using standard ChIP experiments. If confirmed, this could lead on to studies investigating the implications of these interactions on gene expression.

4.2.4 Bioinformatics screen to identify putative gene targets of ZNF350 and IRF-1

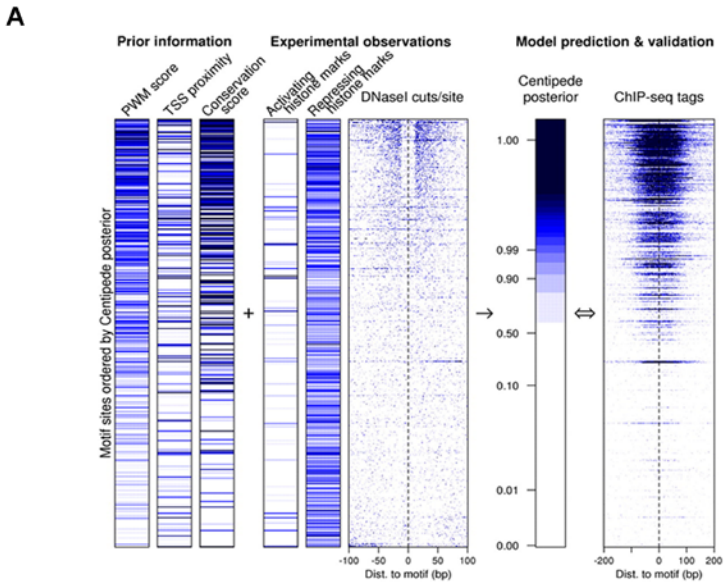
Since the data generated from the ZNF350- and IRF-1- ChIP-seq appears to be quite unreliable due to the lack of overlap between replicates, I sought to use another technique to predict and validate the binding of the two proteins to specific regions of the DNA.

A variety of computational methods have been developed to predict transcription factor binding sites ([316; 317] and references in [318]). These methods generally use sequence-specific binding motifs of individual transcription factors, either obtained from databases such as TRANSFAC or JASPAR [319; 320], or estimated de novo. However, only a small portion of genomic locations containing these motifs actually bind transcription factors. A study by Pique-Regi et al. (2011), has developed an algorithm named CENTIPEDE, that combines cell-specific experimental data with genome sequence information to map bound transcription factor binding sites in a specific sample [318]. Previous work has shown that several genome-wide assays correlate with transcription factor binding, including chromatin accessibility as determined by DNase I sensitivity or FAIRE (formaldehyde-assisted isolation of regulatory elements) assays [321; 322], protection of the transcription factor binding site from cleavage by DNase I [323-325] and ChIP-seq of specific combinations of histone modifications [326] or the co-activator protein, p300 [327]. Additionally, it is known that each type of protein-DNA interaction can produce a characteristic DNaseI footprint that reflects properties of that specific interaction [328]. The CENTIPEDE algorithm integrates these sources of information and importantly correlates very closely with empirical ChIP-seq measurements of transcription factor binding at candidate motif sites [318]. In contrast to ChIP-seq, CENTIPEDE, like other computations methods for inferring transcription factor binding sites, requires the presence of a DNA sequence motif at the binding site [318]. Figure 4-16A illustrates an overview of the CENTIPEDE approach using the transcription factor REST as an example, compared to REST ChIP-seq reads for

validation. The CENTIPEDE algorithm was used to map 827,000 transcription factor binding sites in human lymphoblastoid cell lines, which is comprised of sites corresponding to 239 position weight matrices of known transcription factor binding motifs, and 49 novel sequence motifs [318]. The characterisation of IRF-1 transcription factor binding sites in the CENTIPEDE study generated 7,859 predicted sites throughout the genome. This enabled me to develop a novel computational method for identifying genome locations that posses both IRF-1 and ZNF350 binding motifs and verify these sites using experimental techniques. Figure 4-16B demonstrates the general workflow of identifying these IRF-1 and ZNF350 gene targets.

Briefly, using the web-based Galaxy platform I assembled the predicted IRF-1 binding sites from CENTIPEDE and extracted genomic DNA surrounding the binding site intervals. I was then able to search the genomic sequences for the ZNF350 consensus motif, GGGxxxCAGxxxTTT, within 1 kb of the IRF-1 predicted sites. These results were then filtered for motifs located in promoter regions of genes, since this is thought to be the region transcription factors assemble to enable the regulation of gene expression. I categorised the binding sites up to 10 kb upstream of the TSS due to the possibility that the transcription factors could also be located in enhancer elements several thousand base pairs away from the promoter [329]. After identifying potential gene targets of IRF-1 and ZNF350 by computational methods, I verified the protein-DNA interaction experimentally. By the immunoprecipitation of ZNF350- or IRF-1-bound chromatin, the presence of protein binding on target genes was detected by designing primers around the predicted binding region and analysing gene amplification by PCR.

Using this method, 23 gene targets were identified during the initial computational analysis that contain possible IRF-1 and ZNF350 motifs within close proximity to one another (Table 4-11). The majority of these motifs (approximately 70%) were located within 500 bp upstream of the transcription start site (TSS; Figure 4-17A).



Pique-Regi R et al. Genome Res. 2011;21:447-455

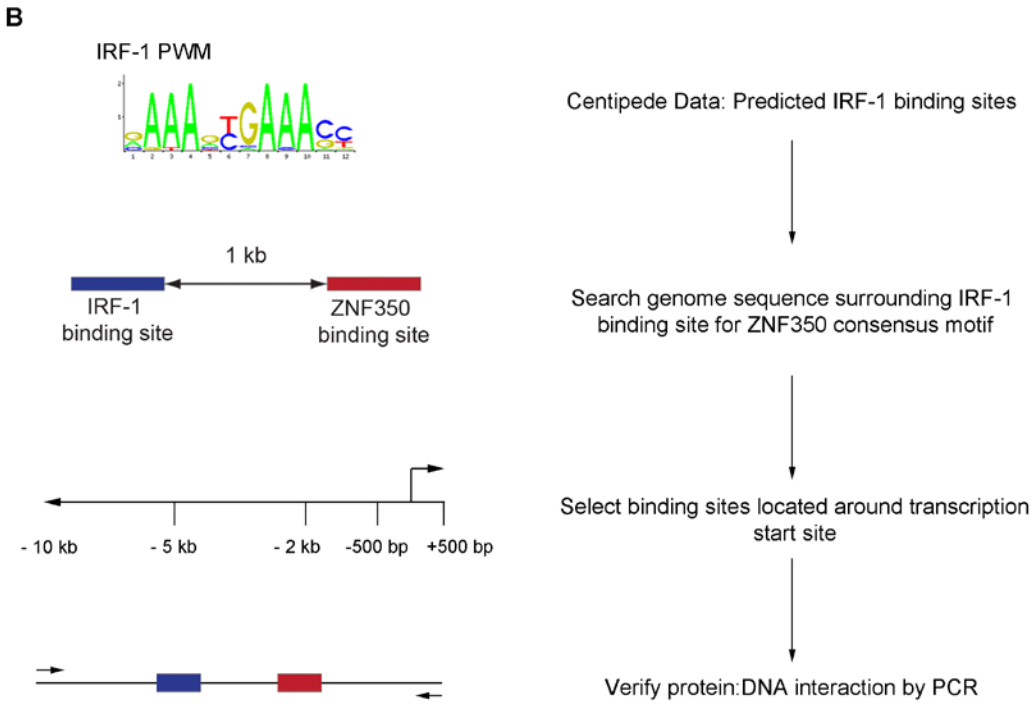


Figure 4-16: ZNF350 and IRF-1 genome binding site prediction using a combination of computational and experimental methods

(A) Overview of the CENTIPEDE approach using the factor REST as an example. Each row in the image represents a genomic location that matches the primary REST binding motif. For each motif instance, prior information (PWM score, TSS proximity, and conservation score) and experimental data (histone marks and DNase-seq reads) was extracted. Rows are ordered by the posterior probability given by the model (bound sites at the top) and, for validation only, compared to REST ChIP-seq reads extracted in a 400-bp window surrounding the motif (last column). Darker blue colouring indicates, in each column, respectively, higher PWM score; motif closer to the TSS; motif site more conserved; more histone ChIP-seq reads in 400-bp windows around each site; more DNase I cuts per site; higher CENTIPEDE posterior probability of binding; larger number of REST ChIP-seq reads per site. The plot shows 200 randomly selected motif sites with posterior probability > 0.5, and 200 with posterior < 0.5, respectively; this sampling procedure increases the fraction of bound sites displayed [318]. (B) Overview of method for prediction of IRF-1 and ZNF350 binding sites using centipede data and the consensus binding motif, respectively.

To confirm the binding of the two proteins *in vivo*, primers were designed around the motif locations and PCR was performed using chromatin immunoprecipitated by anti-ZNF350, anti-IRF-1 or anti-IgG control antibodies alongside an input control to detect amplification of the gene of interest. Of the 23 originally predicted targets, 18 were confirmed to bind one or both proteins (Figure 4-17B and Table 4-11). It can be seen that both ZNF350 and IRF-1 bind to 16 of these genes, whilst only IRF-1 binding is detected on ZNF503 and only ZNF350 interacts with CKS2 (Figure 4-17B). Of the IRF-1 and ZNF350 gene targets identified using this approach several can be classified into groups by their gene function with four, namely GBP3, ID2, IKBKG and IL12A, being involved in the immune response. Interestingly, IRF-1 is thought to be important for the upregulation of interleukin-12 (IL-12), the gene product of IL-12A, in response to LPS in macrophages with a loss of IRF-1 resulting in a considerable reduction of IL-12 inducibility by LPS [266]. In addition, IL-12 has been shown to induce IRF-1 gene expression in human NK and T cells [330], suggesting a possible mechanism of positive feedback regulation. In addition, SP110 has previously been identified as a target of IRF-1 through ChIP-seq in human primary monocytes [273]. It is of interest that this gene also contains a ZNF350 consensus motif however, during the experimental validation of SP110 in this study gene amplification in the input control proved difficult and improved primer design and further analysis will need to be carried out to confirm binding. For the other 4 genes identified to posses ZNF350 and IRF-1 binding sites but were not experimentally validated, the design of suitable primers limited the detection of gene amplification in the input samples. Future studies will be required to identify optimal primers for this analysis. Of the other genes identified using this computational approach, none have been reported to bind either IRF-1 or ZNF350. However, GBP1, a member of the same family of proteins as GBP3, has been identified as an IRF-1 target in the primary monocyte ChIP-seq study [273]. Further experimental studies will be required to elucidate the role of the interaction of both ZNF350 and IRF-1 with these targets in regards to regulation of cellular processes.

Table 4-11: Locations and binding confirmation of predicated ZNF350 and IRF-1 sites

Location from TSS	***Chromo***	***Binding region start***	***Binding region end***	***Associated gene***	***ZNF350 binding***	***IRF-1 binding***
+ 500 bp	chr3	38307710	38309721	SLC22A13	+	+
	chr9	114244557	114246568	KIAA0368	+	++
- 500 bp	chr1	32478249	32480260	KHDRBS1	+	+
	chr1	59610984	59612995	RP11	+	+
	chr1	89487537	89489548	GBP3	+	+
	chr1	146643529	146645540	PRKAB2	+	++
	chr2	8820557	8822568	ID2	-	-
	chr2	220251359	220253376	DNPEP	-	-
	chr2	231090128	231092139	SP110	-	-
	chr3	159705626	159707637	IL12A	+	+
	chr7	151038567	151040578	NUB1	-	-
	chr8	56685410	56687421	TGS1	+	+
	chr9	91925178	91927189	CKS2	+	-
	chr12	104849248	104851259	CHST11	+	+
	chr12	113859005	113861016	SDSL	+	++
	chr15	48624052	48626063	DUT	-	-
	chr19	39420255	39422266	SARS2	++	+
	chrX	153769899	153771910	IKBKG	+	++
-2 kb	chr5	137416104	137418115	WNT8A	+	++
	chrX	118889364	118891375	ANKRD58	-	-
-5 kb	chr10	77156296	77158307	ZNF503	-	+
	chr10	105000740	105002751	LOC729020	-	-
-10 kb	chr10	121127706	121129717	GRK5	+	+

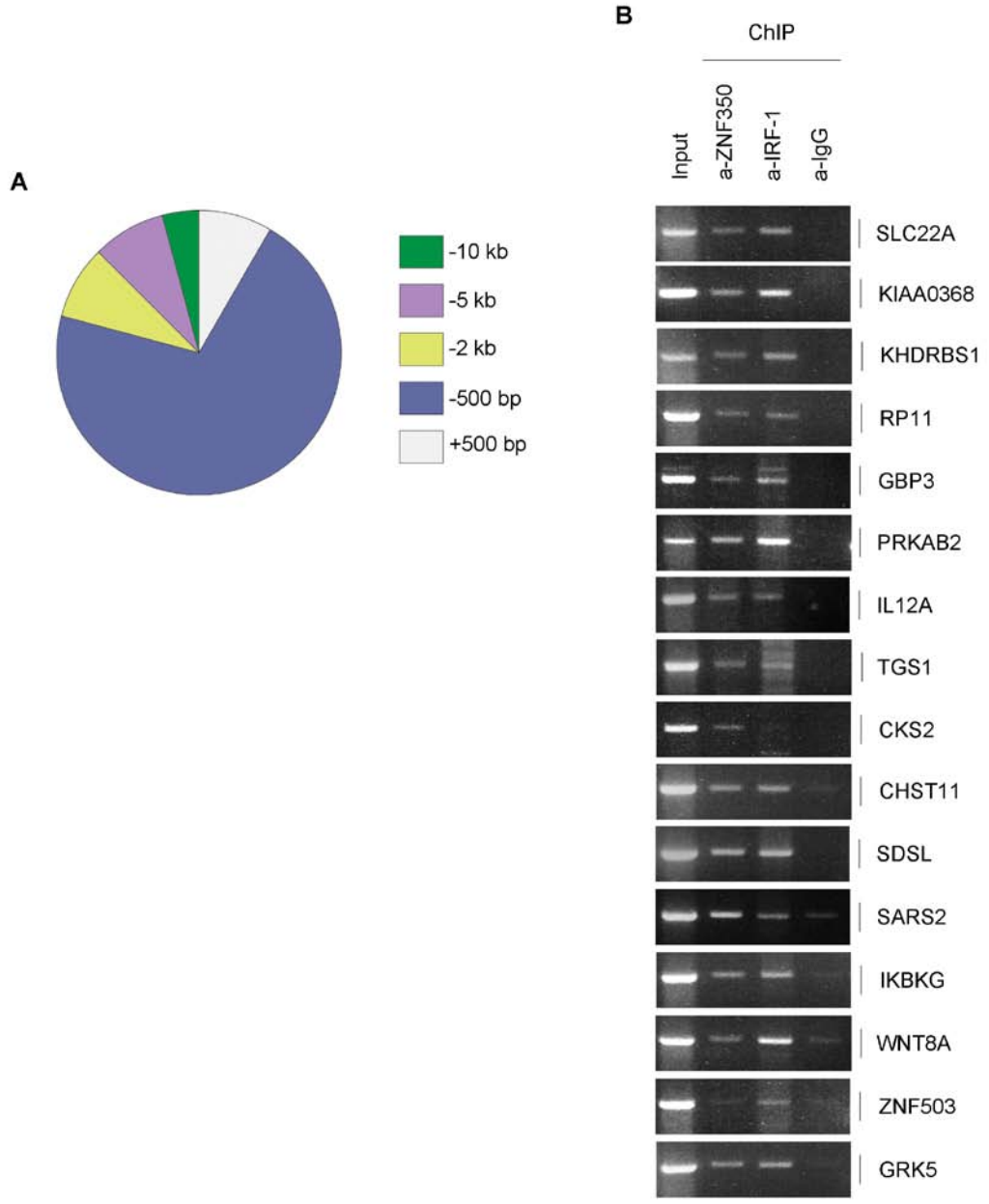


Figure 4-17: Validation of predicted ZNF350 and IRF-1 binding sites

(A) Pie chart showing the distribution of parallel ZNF350 and IRF-1 predicted binding sites in relation to TSS. (B) Agarose gels depicting the amplification of the specified gene. IPs were performed using anti-ZNF350 (Abcam) anti-IRF-1 (H-205) anti-IgG antibodies and PCR was performed along with Input samples using primers designed around the predicted binding site. Primers are described in Table 2.4.

The successful identification of potential transcription factor gene targets followed by experimental validation generates a valuable tool for genome-wide studies of gene regulation. The data presented here successfully demonstrates that IRF-1 and ZNF350 binding sites can be predicted within the genome and validated experimentally. In this study, the parameters of the computational analysis were quite stringent by the use of the specific ZNF350 consensus motif to identify possible binding sites. Of course in reality, ZNF350 may bind to genomic sites that are not as sequence-specific. Indeed, it is possible to identify motifs with nucleotide mismatches using this bioinformatics approach, however the number of potential target genes increases considerably. For example, for predicted IRF-1 sites within 500 bp up- or down-stream of the TSS there are 342 sites with a nearby ZNF350 consensus motif containing 1 mismatched nucleotide, 3445 sites with 2 mismatched nucleotides and 21,021 sites with 3 mismatched nucleotides (data not shown). For the purposes of having a manageable sample size to experimentally validate, I used quite stringent sequence specificity. However, had the ChIP-seq data been of adequate quality, it would have been of great interest to compare the lists of computationally-generated predicted binding sites with the experimentally obtained ChIP-seq binding sites for both ZNF350 and IRF-1. Tentative analysis did not reveal any specific overlap between the ChIP-seq-identified genes and the genes identified by the bioinformatics screen. However, there did appear to be some genes from the same family but of different subtypes and further analysis would be required to explore this. By applying the CENTIPEDE binding site prediction method to ZNF350 consensus sites, the potential number of gene targets could be refined to give a more specific prediction taking into account mismatches of the consensus sequence. This could allow for the validation of a greater number of IRF-1 and ZNF350 binding locations, providing a better insight into the regulation of expression.

4.3 Discussion

Gene-specific regulation of expression is essential for the proper functioning of biological processes such as proliferation, differentiation, development and apoptosis and the dysregulation of gene expression can result in various diseases and disorders. The sequencing of the human genome and its annotation using various comparative genomics and computational methods has resulted in the classification of approximately 20,000-25,000 protein-coding genes [331]. How these genes are regulated and expressed, as well as the effect this has on product function in a cellular environment, are key areas of investigation to enable a better understanding of various biological processes. An important step of this is to identify the transcriptional regulatory elements associated with each gene and under what conditions these may be altered, for example in response to viral infection or DNA damage. Eukaryotic gene transcription is known to be controlled by epigenetic changes and chromatin reorganization, which are mediated by sequence-specific transcription factor binding to proximal or distal regulatory elements [256; 257]. Previous studies have shown that ZNF350 can bind directly to a variety genes and repress their activity. ZNF350 binds to *GADD45A* in a BRCA1-dependent manner to repress expression [45] while a complex of KAP1/LCoR/ZNF350 also binds to *GADD45A* and loss of complex binding results in increased apoptotic cell death of malignant and non-malignant breast epithelial cells [51]. ZNF350 has also been show the repress the expression of *ANG1* [52], *MMP9* [59] and *p21* [55] among others. In a number of single-gene studies, IRF-1 has been identified to bind a number of target genes. IRF-1 is commonly known as a positive regulator of gene transcription with roles in the immune response by targeting genes such as *TAP-1* [267] and *IL-12* [266] as well as DNA damage response genes *p21* and *caspase-1* and apoptotic genes including *caspase-8*, *PUMA*, *TRAIL* and *XAF1* [268-271]. Conversely, IRF-1 has been found to repress the Secretory leukocyte protease inhibitor (*SLP1*) gene, which plays a role in neutrophil-mediated inflammation and the LPS response [123]. In addition, IRF-1 represses *Cdk2*, a component of cell cycle control machinery, with loss of IRF-1 binding to the gene resulting in gene

activation and loss of growth suppression [99]. A ChIP-seq study carried out in human primary monocytes also identified 52 IRF-1-bound regions of genes such as *GBP1*, *TLR3* and *SP110* [273]. To date, only a small number of genes have been identified as being targets of either ZNF350- or IRF-1-mediated repression.

Using sucrose density gradient ultracentrifugation, I was able to show that ZNF350 tightly associates with chromatin *in vivo* (Figure 4-1C) which allowed me to move forward in identifying novel binding sites within the genome. ChIP-seq is a powerful method for the identification of DNA sequences bound by specific proteins. Since there is little data on the way the ZNF350 protein behaves *in vitro* and *in vivo,* I set out to optimize the ChIP to ensure maximum efficiency and success. This involved examining various protocols to identify the best method of ZNF350 and IRF-1 protein extraction from A375 cells and analysing different lysis buffers to ensure optimal protein:DNA levels were used for immunoprecipitation (Figure 4-3). Next, the chromatin must be of a certain size (200-800 bp) for library preparation so I compared different shearing devices (Soniprep 150 and Bioruptor) and optimized the number of sonication pulses required (Figure 4-4). At this stage, I used a bioanalyser (Agilent) to ensure the DNA size was accurately detected since running samples on agarose gels did not appear to be sensitive enough (Figure 4-4 and 4-5). Extensive optimization was carried out to ensure the most favourable conditions for protein-bound DNA analysis. I found that 1% SDS was required for effective sonication of the samples; however this had to be diluted to 0.1% SDS for efficient immunoprecipitation as high levels of SDS interfered with the ChIP signal. The quality of ChIP is determined by the specificity of the antibody and the degree of enrichment achieved in the affinity precipitation step [286]. In order to select the most efficient antibody, I compared different antibodies against ZNF350 and IRF-1 and determined the differences in amplification of known target genes (Figure 4-6). At the time of this study, only two ZNF350 antibodies were suitable for use in immunoprecipitation with one being proposed to be of ChIP-grade quality. Indeed, when comparing *ANG1* gene amplification in each IP, I found that the ChIP-grade anti-ZNF350 antibody (Abcam) performed better than the non-ChIP-grade antibody (Sigma) with clear *ANG1* enrichment (Figure 4-6A) and so this was chosen. Of the

two IRF-1 antibodies analysed, the polyclonal H-205 (Santa Cruz) antibody showed greater *TLR3* amplification compared to the anti-IRF-1 monoclonal antibody (BD) (Figure 4-6B). In addition, the H-205 anti-IRF-1 antibody had been used in a previous IRF-1 ChIP-seq study in primary monocytes [273] that generated successful results so I decided that to use this IRF-1 antibody.

Following high-throughput sequencing of the protein-bound chromatin and initial analysis of the reads, I discovered that the biological replicates of the two IP conditions did not show any considerable overlap, indicating a problem had occurred during the process. By determining that the actual sequencing of the chromatin was of satisfactory quality, it would indicate that the experimental ChIP had caused the problems with the data. After analysing each step and ensuring the best conditions for ChIP were used, there are no clear indicators of what caused the issues. One possible reason could be the use of polyclonal antibodies since they can recognize several different epitopes on the same protein to achieve a greater amount of sensitivity (i.e. pull-down power), which may sacrifice the specificity of the antibody [300]. However, the use of the same anti-IRF-1 polyclonal antibody (H-205, Santa Cruz) in a previous IRF-1 ChIP-seq study was of good quality and did not appear to cause any issues with their data [273]. As previously stated, at the time of this study there were no commercially available ZNF350 monoclonal antibodies however, recently, one has been developed (Abnova). To ensure successful ChIP-seq results in any future studies, the use of robust monoclonal antibodies should be explored and this may require the development of a novel ZNF350 mAB.

Despite the ChIP-generated data being of unsatisfactory quality to gain any significant insights into the genome-wide association of ZNF350 or IRF-1, I proceeded to compare the biological replicate data for each transcription factor. I found similarities in the position of both transcription factors in relation to TSS, (Figures 4-14 and 4-15) however no significant consensus motif was identified for either ZNF350 or IRF-1 (Tables 4-4, 4-5, 4-8 and 4-9). Due to the low quality of the data, very little overlap was observed between the replicates of each IP condition (Figure 4-12), however of the 10 ZNF350 overlapping sites (Table 4-3) and 9 IRF-1 overlapping sites (Table 4-6), 6 common sited were identified (Table 4-10). These

sites corresponded to 3 genes, namely *MDM2*, *ZNF350* and *IRF-1*. If this finding is a true feature of ZNF350 and IRF-1 binding and not an artefact of the dubious data, it provides interesting implications for the regulation of ZNF350 and IRF-1 gene expression. Providing the ChIP-seq is of good quality, a vast amount of information can be derived from the data produced. This may be the analysis of consensus binding motifs, location of sites throughout the genome such as within promoters, exons, 5'UTR etc., location of sites in relation to chromatin accessibility, histone marks and so on.

As a way to further identify novel ZNF350 and IRF-1 target genes after the ChIP-seq problems, I developed an alternative method involving the initial computational prediction of binding sites followed by experimental validation. This method proved successful in the identification of several genes that possess both a ZNF350 and an IRF-1 binding site with the validation of these interactions in approximately 70% of novel target genes (Figure 4-17). In support of the successful use of this method to identify predicted TFBS, SP110 was identified both using CENTIPEDE and in a previous IRF-1 ChIP-seq study [273]. The CENTIPEDE algorithm integrates various sources of information such as chromatin accessibility as determined by DNaseI sensitivity or FAIRE (formaldehyde-assisted isolation of regulatory elements) assays [321; 322], protection of the transcription factor binding site from cleavage by DNaseI [323-325], ChIP-seq of specific combinations of histone modifications [326] and DNaseI footprints that reflect the properties of a specific interaction [328]. This algorithm has been shown to be comparable to experimental ChIP-seq data of transcription factor binding at candidate motif sites [318]. CENTIPEDE has predicted IRF-1 binding sites within the genome which enabled me to analyse the presence of proximal ZNF350 sites. The use of CENTIPEDE to predict ZNF350 binding sites would be of great use, however the algorithm requires the presence of a DNA sequence motif at the binding site [318]. The ZNF350 consensus motif was characterized by the purification of GST-ZNF350 with a random pool of 55-mer oligonucleotides and bound sequences subjected to several rounds of amplification. A final round of 16 subclones was then aligned to derive the GGGxxxCAGxxxTTT motif [43]. To ensure the ZNF350 consensus motif is

accurately defined, a wider analysis of ZNF350 biding sites should be analysed, perhaps that from a successful ChIP-seq experiment.

Despite the ChIP-seq data not being useful for the identification of novel ZNF350 and IRF-1 binding sites, I was able to develop a new method using computational tools followed by experimental validation to identify 16 novel target genes containing binding sites for both transcription factors. This information can be used in future investigations to understand the mechanisms of ZNF350 and IRF-1 gene regulation and what consequences this may have on biological processes in a cellular environment.

Chapter 5: Peptide phage display combined with High-throughput Sequencing to identify ZNF350-interacting peptides

5.1 Introduction

Peptide phage display is a widely-used method for the identification of protein ligands. Since its discovery, phage display has influenced many scientific fields including (i) proteomics, where it has been used to analyse protein-protein interactions and identify novel enzyme substrates and inhibitors ([332], and references in [333]), and (ii) drug discovery and design with the development of vaccines and the discovery of agents for targeted delivery of drugs and gene therapy ([334; 335] and references in [333]). The technology was first described by George Smith in 1985, who reported that foreign peptides could be expressed on the surface of filamentous bacteriophage, which is achieved by inserting a peptide-encoding gene into the gene for a capsid structural protein [336]. This initial discovery has been developed over the years to generate phage-displayed peptide libraries, with the ability to screen as many as 10^{10} different peptides simultaneously [337; 338]. Specifically, 7-mer, 12-mer, 15-mer or n-mer peptides (20^n, where 20 is the number of standard amino acids and n is the number of randomized positions) are genetically fused to any of the five capsid proteins from the phage (pIII, pVI, pVII, pVIII and pIX), resulting in billions of pooled peptides being presented on phage particles [337-339]. These libraries are then used to identify a phage that specifically binds to a target protein. Following rigorous washing to remove non-specific phage, the target-bound phage is propagated by amplification in *E.coli* and purified to create a pool of phage containing only peptides that display an affinity for the target protein. This pool can be used in additional rounds of affinity selection in a process known as

"biopanning". After multiple rounds of biopanning, the peptide-displayed phage library is gradually enriched for high-affinity binders [340; 341]. Traditionally, peptide sequences are analysed by single clone amplification followed by Sanger sequencing, however this method is quite labour intensive and is limited to analysis of a maximum of one hundred clones, usually one by one.

By analysing a small number of sequences enriched in a screen, a consensus binding motif can be predicted for the target of interest. However, phage display screens may also identify a large number of consensus motifs. This may be the case for a transcription factor that is involved in many, transient interactions with a wide variety of different target proteins. If a screen selects a large number of independent binding clones, they all must be sequenced to identify all the possible consensus motifs [342]. Dias-Neto et al. (2009) were the first group to adapt next-generation pyrosequencing (454/Roche) to analyze ~50,000 sequences from a library of 7-mer peptides [343]. Subsequently, the same sequencing technique has been used by other groups including Ernst et al. (2010), who used barcoded primers for the preparation of a library to enable them to sequence 22 independent panning experiments in one run [344]. Higher throughput sequencing technologies such as Illumina/Solexa could provide more complete coverage of the libraries and enable the analysis of multiple experiments in parallel. Indeed, Illumina sequencing has been utilised to characterize phage-displayed libraries of single chain antibodies (scFv) [345] while a separate study has shown that deep sequencing of the phage pool after just one round of biopanning is sufficient to identify positive phages against cell surface receptors [346].

Since phage display technology has been used in over 5000 publications and patents in the past 20 years [342], characterization of sequence diversity in phage libraries is fundamentally important. Phage display has been used to identify ligands for hundreds of targets, yet several observations are poorly explained. For example, identical sequences can emerge from unrelated screens for unrelated targets [347], and screens that should present a large number of diverse ligands actually only produce one sequence motif (reviewed in [348]). Recently, a study has used deep-sequencing to characterize the diversity in phage libraries. The authors found that

screens based on repeated amplification and traditional small-scale sequencing only identify a few binding clones and miss thousands of useful clones [342]. Therefore, by developing new screening strategies and utilising the deep sequencing approach, the diversity of phage clones can be preserved and ligands previously lost in conventional phage display screens may be identified.

ZNF350 was initially identified as an IRF-1-interacting protein through phage display technology [160]. The work in this thesis has verified this physical interaction and demonstrated the activity of the two proteins as transcriptional regulators. Transcription factors tend to be involved in many transient interactions to enable the regulation of gene expression. Currently, ZNF350 is known to interact with IRF-1 [160] and BRCA1 [45], however since transcriptional regulators often act in complex with many different proteins, I sought to identify novel ligands of ZNF350 to facilitate better understanding of the repressor complex ZNF350 may be involved in. This chapter details the use of peptide phage display combined with next-generation sequencing to identify peptide sequences that bind to ZNF350.

5.2 Results

Phage display was carried out as detailed in Figure 5-1, where recombinant His-ZNF350 was immobilised on an ELISA plate, followed by the addition of the Ph.D.-12 Phage display peptide library (NEB). This library contains approximately 10^9 distinct clones consisting of linear 12-mer peptides fused to the N-terminus of the pIII minor coat protein of M13 phage, with a short glycine-serine spacer (Gly-Gly-Gly-Ser) in between. Two rounds of panning were conducted with two types of washes: (i) six fast washes where the wash buffer was added then immediately discarded, or (ii) six slow washes where the wash buffer was incubated on the plate for 5 min each. The phages were then eluted and amplified by infection of ER2378 cells followed by precipitation of phage particles. After a second round of biopanning, PCR was performed using the eluted phage to amplify the peptide DNA sequence. The primers used in the PCR reaction were designed specifically for use in Illumina HiSeq2000 multiplex paired-end read sequencing and comprised of an adaptor sequence, a sequencing key specific for each sample and the phage DNA template sequence.

5.2.1 Sequencing quality filters

Initial processing of the phage display sequencing data revealed 8.2 million sequences, 2.4 million of which were unique. The amplification of phage libraries results in the decreased diversity of sequences which can limit the number of binding clones identified in the screen [348]. The library used in this study showed good quality diversity with most sequences showing just 1 copy with the maximal number of copies being around 100 (Figure 5-2A). Poorly diverse libraries tend to show a much higher number of copies per library (Figure 5-2B).

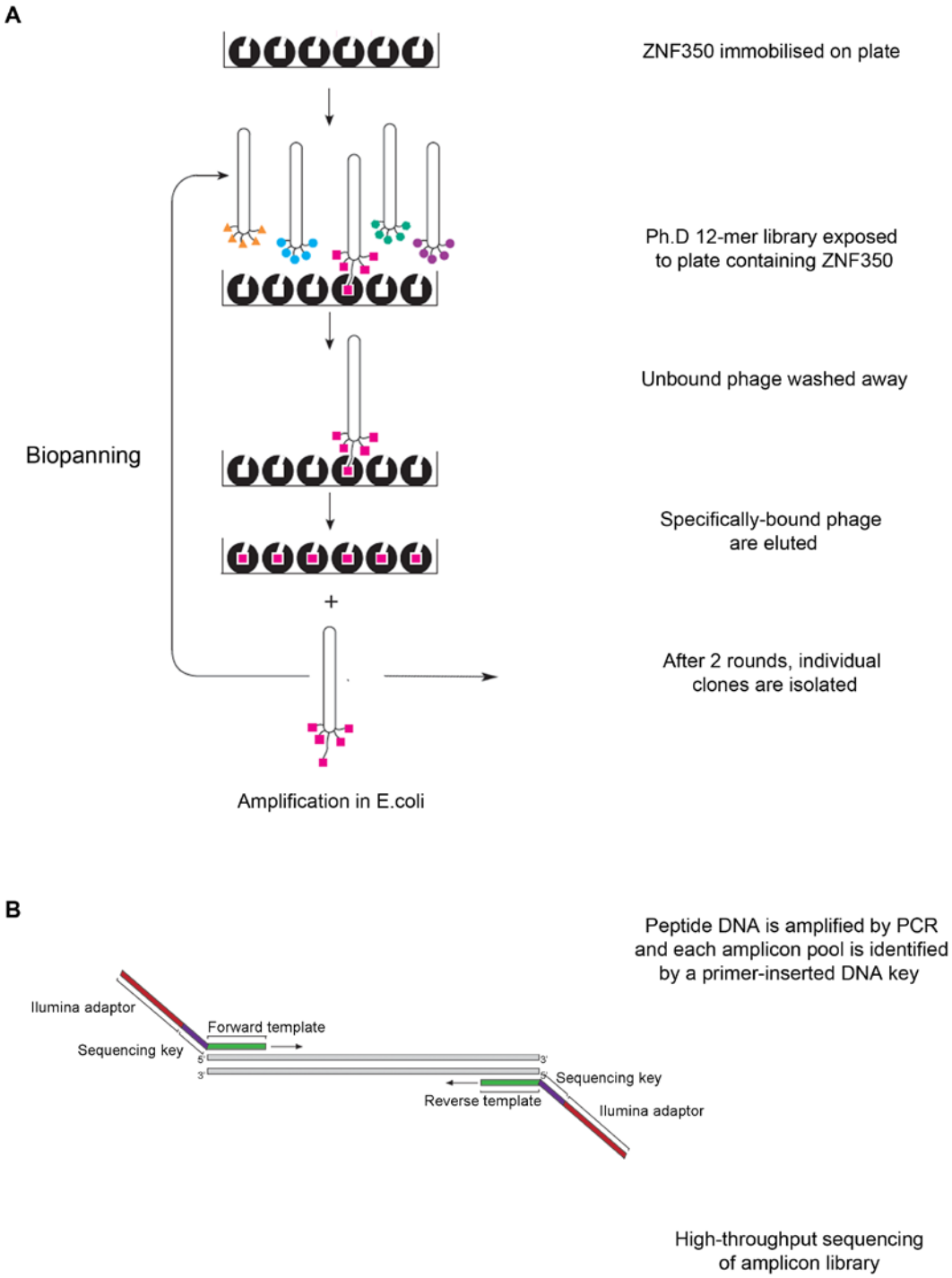


Figure 5-1: Phage display biopanning approach using next-generation sequencing

(A) Panning with a Ph.D Phage Display Library. Image modified from NEB figure [349]. (B) Design of primers for phage insert amplification by PCR, followed by high-throughput sequencing using Illumina HiSeq 2000.

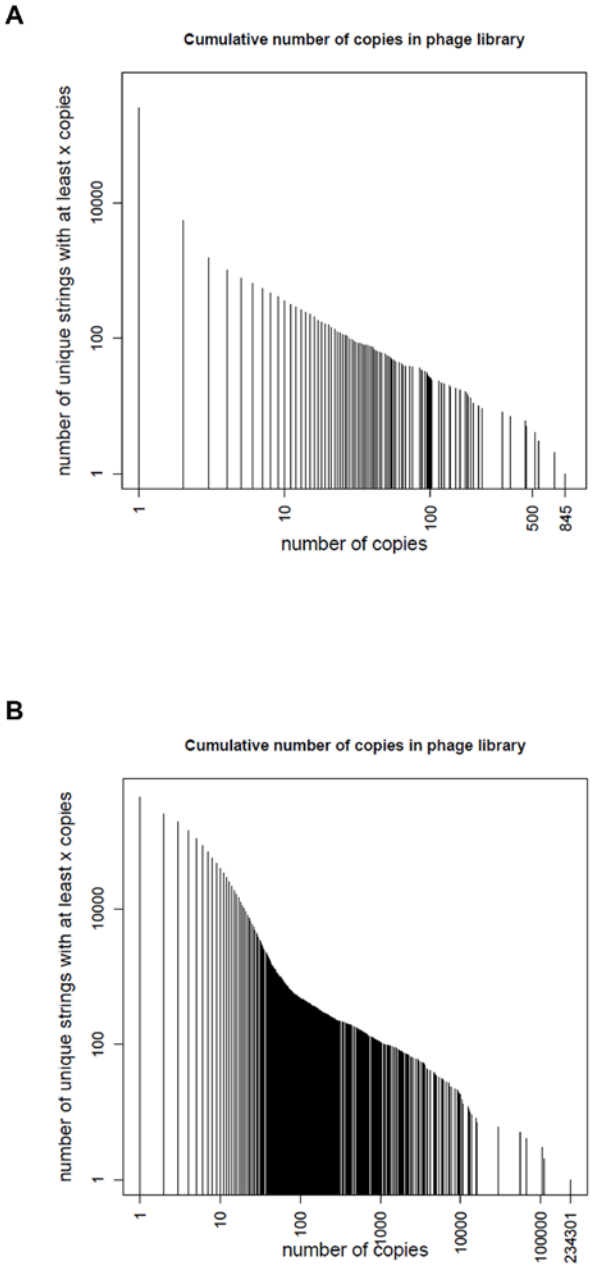


Figure 5-2: Diversity of phage display library

(A) Graph showing cumulative number of copies in phage library used in this study. (B) Example of a poorly diverse library. Both images were generated by Adam Krejčí.

The first step in data analysis is to extract each read with the identification code corresponding to the peptide sequence. Data extraction was processed in collaboration with Stuart Aitken (Semple Group) who generated a Java program to identify the peptides and count the number of reads in which they occur. The reads were filtered for having the correct sequence of nucleotides: identification code>CGC...TCT<peptide>GGT. This criterion ensures that the insert has a proper 36-nucleotide sequence. The sequencing quality scoring system with Illumina is measured by the Phred quality score (Q score), which assesses the base calling accuracy. It indicates the probability that a given base is called incorrectly by the sequencer. The reads in this study were filtered to have a Phred score of 53 or above, which provides a base call accuracy of 99.999% [350]. The Java program outputs an excel table containing these quality filtered sequences and is truncated when the number of reads per peptide is 3 or fewer to account for the memory limit of an excel file. Table 5-1 shows the total number of reads generated for the six different ZNF350 panning conditions along with the actual number of analysed reads. When phage display panning was carried out with fast washes, after the second round there is a reduction in sequenced reads. This could be due to different selection pressures, for example it is known that some sequences in phage libraries amplify slower in bacteria than others which may lead to a progressive collapse in diversity and a smaller number of binding sequences [351]. The commercial 12-mer phage library was also sequenced along with the samples in order to identify sequences that may be over-represented (second column).

The phage display was carried out as part of a multi-protein experiment in collaboration with other colleagues analysing NPM, synuclein and PDE4, along with a no protein control. This resulted in a total of 39 different conditions sequenced in parallel, giving a high number of controls to be used during analysis.

Table 5-1: Number of sequenced reads per sample

Sequence code	***AAA***	***TGT***	***TGG***	***TGC***	***TCA***	***TCT***	***TCG***
Sample	12-mer library	ZNF350 Fast wash Round 1 amplified	ZNF350 Fast wash Round 1 eluted	ZNF350 Fast wash Round 2 amplified	ZNF350 Slow wash Round 1 amplified	ZNF350 Slow wash Round 1 eluted	ZNF350 Slow wash Round 2 amplified
total number of reads	6,637,832	180,838	172,697	467	185,510	184,050	265,464
number of analysed reads	4,017,778	141,671	19,358	332	155,960	21,244	221,737

5.2.2 Analysis of next-generation sequencing data

Since next-generation sequencing of enriched peptide sequences in a phage display screen is a relatively new process, there is no apparent format of analysis to determine how significant enrichment is. Therefore, I sought to establish a method for analysing and identifying significantly enriched peptides of ZNF350 by utilising the large number of controls generated from the collaborative screen. Due to the very low number of reads sequenced for the 2nd round of panning with fast washes (Table 5-1), I decided to focus on the reads collected after slow washes to identify specific enrichment after multiple rounds of biopanning. The Java program outputs a large excel file containing a list of all the sequences, the sum of all the reads per sequence and the number of reads per condition. First, I calculated the number of reads for ZNF350 slow wash round 1 amplified (ZNF350 S1) and round 2 amplified (ZNF350 S2) combined and worked this out as a percentage of the sum of all reads (39 conditions plus the library). This list could then be sorted by percentage of total

reads, Table 5-2 lists the top 5 peptides when analysed this way. Indeed for 170 peptides, ZNF350 possess 100% of all the reads per sequence. It can also be seen that there were no peptide reads present for the library, indicating that those identified for the ZNF350 S1 and S2 conditions are specific and not generally over-represented. However despite being very specific, there are often a low number of reads per sequence, which is not ideal as we want to look at highly enriched sequences corresponding to a larger number of reads.

Table 5-2: Reads sorted by ZNF350 percentage of sum

peptide	*sum of all reads*	*12-mer library*	*ZNF350 S1*	*ZNF350 S2*	*Total (S1+S2)*	*Percentage of sum*
KPVADRWSYAPK	21	0	15	6	21	100%
NNSKTMDIAEAD	21	0	12	9	21	100%
TGWVKHMVSLPR	21	0	3	18	21	100%
ELMANMPETACP	21	0	2	19	21	100%
TGNENLPRVQAI	20	0	10	10	20	100%

To circumvent this, I filtered the list for ZNF350 reads to be a minimum of 70% of the sum of all reads and then sorted the data by the total reads of ZNF350 S1 + S2 (Table 5-3, showing top 5 peptides). As a result of this criterion, the number of distinct sequences decreased to 4,675. These remaining peptides are therefore specific of ZNF350 and in order of enrichment after two rounds of amplification

Table 5-3: Reads sorted by ZNF350 total with minimum 70% of sum

peptide	*sum of all reads*	*12-mer library*	*ZNF350 S1*	*ZNF350 S2*	*Total (S1+S2)*	*Percentage of sum*
NPQPSPMGSRAL	6760	29	3063	2404	5467	81%
GWYAASGTSLLS	5106	20	1274	2555	3829	75%
LPLVPYQYHLRI	3414	13	951	1687	2638	77%
SAAYAVGDVHRE	3015	10	1151	1136	2287	76%
FPMSYDWNHDWL	2694	10	521	1607	2128	79%

To confirm that the peptides with the most reads for ZNF350 are significant, I calculated the fold change between the sample and the control. This is done by dividing, for example, ZNF350 S1 read number by either library (Table 5-4, showing top 5 peptides) or control S1 (Table 5-5, showing top 5 peptides) read number. Since dividing by 0 is impossible, iteration numbers of controls with such value were incremented to 1. If the fold change for a given sequence was less than 5-fold in any round, it was not considered to be significant. This choice of fold increase was arbitrary and could be increased or decreased to a more standard 2-fold, however it is a balance between loss of information and the risk of having false positives. By filtering the data initially by percentage of total reads then eliminating sequences with a lower fold change, the list of remaining ZNF350 peptides is ensured to be highly significant.

Table 5-4: Fold change between ZNF350 and library reads

peptide	*sum*	*12-mer library*	*ZNF350 S1*	*ZNF350 S2*	*Fold change S1*	*Fold change S2*
NPQPSPMGSRAL	6760	29	3063	2404	105.62	82.90
GWYAASGTSLLS	5106	20	1274	2555	63.70	127.75
LPLVPYQYHLRI	3414	13	951	1687	73.15	129.77
SAAYAVGDVHRE	3015	10	1151	1136	115.10	113.60
FPMSYDWNHDWL	2694	10	521	1607	52.10	160.70

Table 5-5: Fold change between ZNF350 and Control reads

peptide	*sum*	*Control S1*	*Control S2*	*ZNF350 S1*	*ZNF350 S2*	*Fold change S1*	*Fold change S2*
NPQPSPMGSRAL	6760	64	55	3063	2404	47.86	43.71
GWYAASGTSLLS	5106	38	35	1274	2555	33.53	73.00
LPLVPYQYHLRI	3414	25	16	951	1687	38.04	105.44
SAAYAVGDVHRE	3015	23	23	1151	1136	50.04	49.39
FPMSYDWNHDWL	2694	15	15	521	1607	34.73	107.13

5.2.2.1 Identification and analysis of enriched motifs

This list of peptides was used to search for linear binding motifs with MEME motif finder [352]. The parameters of the MEME program can be altered, so as a starting point, the minimum motif width was set to 4 residues while the minimum optimum number of sites for each motif was set to 2. Two very distinct motifs were identified through MEME with the most statistically significant being ATMPVSAG[EK] (Figure 5-3A), which was identified in 7 peptides of the enriched pool. This motif was entered into ScanProsite [353], which is a computational tool that can rapidly and reliably identify proteins from the UniProt database containing the specific motif. Variations of the motif were scanned using ScanProsite, with a minimum of 5 conserved residues required for a protein match. Figure 5-3B shows the proteins with the most conserved motifs while Appendix VI lists all identified proteins in detail. Of interest, proteins including histone deacetylase 4 and transcription initiation factor TFIID subunit 10 (TAF10) were identified, which are involved in transcriptional regulatory complexes.

A

Motif ATMPVSAG[EK] E-value 3.90E-21

p-value	Sites	
9.45E-12	ATMPVSAGE	SVS
9.45E-12	ATMPVSAGE	AIF
9.45E-12	ATMPVSAGE	AIS
9.45E-12	ATMPVSAGE	SIF
9.45E-12	ATMPVSAGE	SIS
1.75E-11	ATMPVSAGK	SIF
1.75E-11	ATMPVSAGK	SIS

B

Uniprot ID	Protein ID	Name	Pattern
Q5T011	SZT2	Protein SZT2	A-T-M-P-X-S-A-G
Q9H7T0	CTSRB	Cation channel sperm-associated protein subunit beta	A-T-M-P-X-S-A
Q9NQ94	A1CF	APOBEC1 complementation factor	A-T-X-P-V-S-A
Q3C1V0	M4A18	Membrane-spanning 4-domains subfamily A member 18	A-T-X-P-V-S-A
Q7Z3T8	ZFY16	Zinc finger FYVE domain-containing protein 16	A-T-X-P-V-S-A
Q6ZP82	CC141	Coiled-coil domain-containing protein 141	A-X-M-P-V-S-A
O95359	TACC2	Transforming acidic coiled-coil-containing protein 2	A-X-M-P-V-S-A
P0C2Y1	NBPF7	Putative neuroblastoma breakpoint family member 7	T-M-X-V-S-A-G

C

Uniprot ID	Protein ID	Name	Pattern
P56524	HDAC4	Histone deacetylase 4	A-T-X-P-V-S
Q7Z6Z7	HUWE1	E3 ubiquitin-protein ligase HUWE1	P-V-S-A-G
Q12962	TAF10	Transcription initiation factor TFIID subunit 10	P-V-S-A-G

Figure 5-3: ATMPVSAG[EK] motif identified in ZNF350 binding peptides

(A) ATMPVSAG[EK] identified by MEME [352] as an enriched ZNF350 binding motif. 7 identified peptides and their p values. The E-value (Expect value) is a parameter that describes the number of hits to be expected to see by chance when searching a particular database. The lower the E-value, the more significant the match is. (B) List of proteins containing the most conserved motifs. (C) Notable protein findings containing the motif.

The other statistically significant motif identified through the MEME program was NPQPSPMGS, which was present in 6 enriched peptides (Figure 5-4A). As above, this peptide was entered into the ScanProsite protein identification program and a large list was generated, the most conserved of which are detailed in Figure 5-4B with the full list found in Appendix VII. Of particular significance is the histone deacetylase p300, which is known to interact directly with IRF-1. In addition, 3 other proteins involved in transcription cofactor activity were identified which were DNA-binding protein RFX7, mediator of RNA polymerase II transcription subunit 13 (MED13) and nuclear receptor co-repressor 2 (NCOR2). It will be of particular interest to verify experimentally if any of these proteins directly interact with ZNF350 *in vivo*.

Panther is an online tool that can analyse a list of genes or proteins and functionally class them into various ontologies [354]. The protein lists obtained using both ATMPVSAG[EK] and NPQPSPMGS motifs were uploaded to Panther to analyse the distribution of proteins in regards to molecular function, biological processes (Figure 5-5A and B) and cellular pathways (Figure 5-6). Panther can also analyse and compare classifications of a list of genes or proteins to a reference list (in this case the whole genome), to statistically determine over- or under- represented categories. The most statistically significant over-represented molecular functions are demonstrated in Table 5-6, with all functions being involved in binding, including sequence-specific DNA binding transcription factor activity. These results are particularly interesting due to the role of ZNF350 in transcriptional regulation. Table 5-7 details the statistically significant biological processes identified for ZNF350-binding peptide protein lists with notable processes such as biological regulation, DNA-dependent transcription, regulation of transcription from RNA polymerase II promoter and cell adhesion over-represented. Panther analysis was also carried out for classification and over-representation of cellular pathways for the ZNF350-binding peptide protein list. Figure 5-4 illustrates all the pathways occupied by the proteins analysed while Table 5-8 shows only those that are statistically over-represented when compared to expected results from the entire genome. Interestingly, B-cell activation and T-cell activation pathways are over-represented

highlighting a possible role for ZNF350 in the immune response. In addition, the Wnt-signalling pathway is over-represented with the ZNF350-peptide protein list. This is a conserved pathway in eukaryotes that is involved in regulating important processes such as cell migration, cell polarity, determination of cell fate and organogenesis during embryonic development [355].

A

Motif NPQPSPMGS E-value 2.00E-11

p-value	Sites	
1.57E-11	NPQPSPMGS	LAL
1.57E-11	NPQPSPMGS	REL
1.57E-11	NPQPSPMGS	RAL
2.08E-11	NPQPIPMGS	RAL
3.11E-11	NPQPSPMGA	RAL
1.45E-10	NPHPSPMGS	RAL

B

Uniprot ID	Protein ID	Name	Pattern
Q8N6Y0	USBP1	Usher syndrome type-1C protein-binding protein 1	N-P-Q-P-S-P-X-G-S
Q6WKZ4	RFIP1	Rab11 family-interacting protein 1	N-P-X-P-S-P-M-G
A6ND36	FA83G	Protein FAM83G	N-P-Q-P-S-X-M
Q9C0D5	TANC1	Protein TANC1	N-P-X-P-S-P-M
P41235	HNF4A	Hepatocyte nuclear factor 4-alpha	P-Q-P-S-P-X-G
Q8N196	SIX5	Homeobox protein SIX5	P-Q-P-S-P-X-G
Q9ULT6	ZNRF3	E3 ubiquitin-protein ligase ZNRF3	P-Q-P-S-P-X-G
Q86UP3	ZFHX4	Zinc finger homeobox protein 4	P-Q-P-S-X-M-G
Q16825	PTN21	Tyrosine-protein phosphatase non-receptor type 21	Q-P-S-P-M-X-S
Q9UPR6	ZFR2	Zinc finger RNA-binding protein 2	Q-P-S-P-X-G-S
O60292	SI1L3	Signal-induced proliferation-associated 1-like protein 3	Q-P-S-X-M-G-S

C

Uniprot ID	Protein ID	Name	Pattern
Q09472	EP300	Histone acetyltransferase p300	N-P-Q-P-X-P
Q9UHV7	MED13	Mediator of RNA polymerase II transcription subunit 13	P-S-P-X-G-S
Q9Y618	NCOR2	Nuclear receptor corepressor 2	P-S-X-M-G-S
Q2KHR2	RFX7	DNA-binding protein RFX7	P-S-P-M-X-S

Figure 5-4: NPQPSPMGS motif identified in ZNF350 binding peptides

(A) Identification of NPQPSPMGS as enriched ZNF350-bidning peptide using MEME software [352]. 6 identified peptides and their p values. The E-value is also calculated and displayed by MEME. (B) List of proteins containing the most conserved motifs. (C) Notable protein findings containing the motif.

Figure 5-5: Classification of ZNF350-peptide protein lists

Pie charts compiled in Panther detailing ZNF350-peptide proteins lists by (A) molecular function and (B) biological processes.

Table 5-6: Over-represented molecular functions for ZNF350-peptide protein lists

Molecular Function	***No. of hits from whole genome***	***No. of hits from ZNF350 list***	***expected***	***+/-***	***P value***
DNA binding	2032	54	24.79	+	6.65E-06
sequence-specific DNA binding transcription factor activity	1754	48	21.4	+	1.86E-05
nucleic acid binding transcription factor activity	1762	48	21.5	+	2.13E-05
nucleic acid binding	3315	71	40.44	+	1.57E-04
binding	5933	106	72.38	+	8.57E-04
small GTPase regulator activity	400	17	4.88	+	1.77E-03
ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity	48	5	0.59	+	5.42E-02

Table 5-7: Over-represented biological processes identified for ZNF350-peptide lists

Biological Process	***No. of hits from whole genome***	***No. of hits from ZNF350 list***	***expected***	***+/-***	***P value***
biological regulation	3252	79	39.67	+	1.17E-07
cellular process	5952	114	72.61	+	6.21E-06
transcription, DNA-dependent	1987	53	24.24	+	8.99E-06
transcription from RNA polymerase II promoter	1976	52	24.11	+	1.89E-05
RNA metabolic process	2473	59	30.17	+	6.26E-05
regulation of nucleobase-contain. compound metabolic process	1648	44	20.1	+	1.48E-04
regulation of transcription from RNA polymerase II promoter	1546	41	18.86	+	4.37E-04
developmental process	2846	61	34.72	+	1.28E-03
regulation of biological process	2175	50	26.53	+	1.67E-03
pattern specification process	244	13	2.98	+	2.18E-03
cell adhesion	890	27	10.86	+	2.77E-03
biological adhesion	890	27	10.86	+	2.77E-03
cellular component organization	1065	30	12.99	+	3.71E-03
cellular component organization or biogenesis	1137	30	13.87	+	1.21E-02
nucleobase-containing compound metabolic process	3532	67	43.09	+	2.02E-02
dorsal/ventral axis specification	39	5	0.48	+	2.32E-02
anatomical structure morphogenesis	691	21	8.43	+	2.45E-02
cell communication	3221	62	39.29	+	2.63E-02
ectoderm development	909	25	11.09	+	2.69E-02

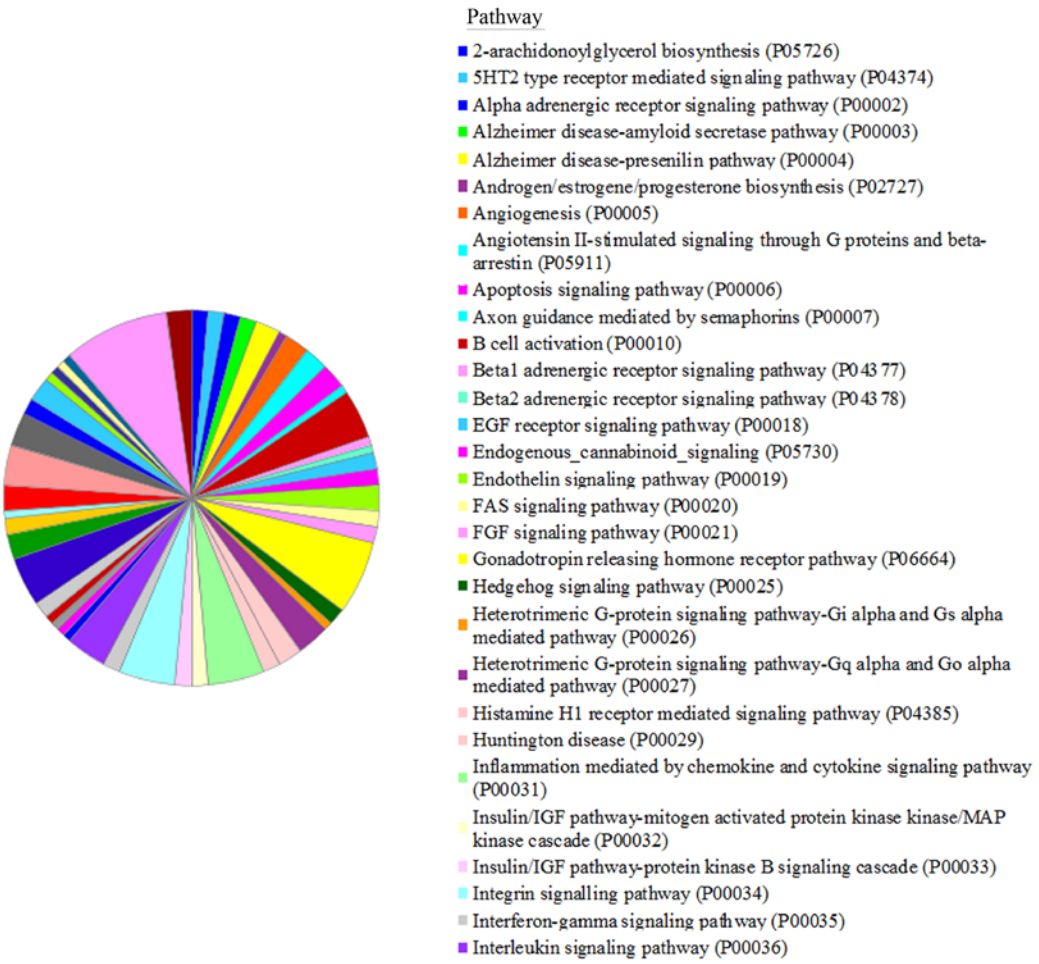


Figure 5-6: Classification of ZNF350-peptide protein lists into pathways
Panther pie charts of ZNF350-peptide proteins lists by pathways

Table 5-8: Over-represented pathways identified for ZNF350-peptide protein lists

Pathways	***No. of hits from whole genome***	***No. of hits from ZNF350 list***	***expected***	***+/-***	***P value***
B cell activation	64	7	0.78	+	2.96E-03
Wnt signaling pathway	297	14	3.62	+	3.94E-03
Gonadotropin releasing hormone receptor pathway	228	11	2.78	+	2.44E-02
T cell activation	82	6	1	+	1.01E-01

From the top 10 motifs identified through MEME, 5 hits contained two cysteine residues in various configurations (Figure 5-7). Cysteine residues are known to display high affinity towards zinc ions, with these resulting Zn^{2+}-cysteine complexes being critical mediators of protein structure, catalysis and regulation [356]. ZNF350 consists of 8 C2H2 zinc fingers so it could be hypothesised that these display an affinity toward cysteine residues present in other proteins to facilitate binding and regulation. An interesting example of this is the interaction between the zinc finger protein Nizp1 and the SET-domain histone lysine methyltransferase NSD1. Nizp1 contains multiple C2H2 zinc finger motifs and a unique C2HR motif, which functions to mediate a protein-protein interaction with the cysteine-rich (C5HCH) domain of NSD1, in a Zn^{2+}-dependent manner. Nizp1 has been shown to repress transcription in a NSD1-dependent manner and mutation of the zinc-coordinating cysteine residues in the C2HR motif of Nizp1 abolishes the interaction with NSD1 and compromises transcriptional repression [357]. Cysteine residues in zinc fingers may be spaced a varying intervals, for example C2H2 zinc fingers take the conformation X_2-Cys-$X_{2,4}$-Cys-X_{12}-His-$X_{3,4,5}$-His [28] whereas RING finger domains, another type of zinc finger, have the consensus sequence Cys-X_2-Cys-$X_{[9-39]}$-Cys-$X_{[1-3]}$-His-$X_{[2-3]}$-Cys-X_2-Cys-$X_{[4-48]}$-Cys-X_2-Cys [358]. With the identification of an abundance of cysteine-rich motifs in ZNF350 binding peptides, it could highlight a possible mechanism by which ZNF350 regulates transcription.

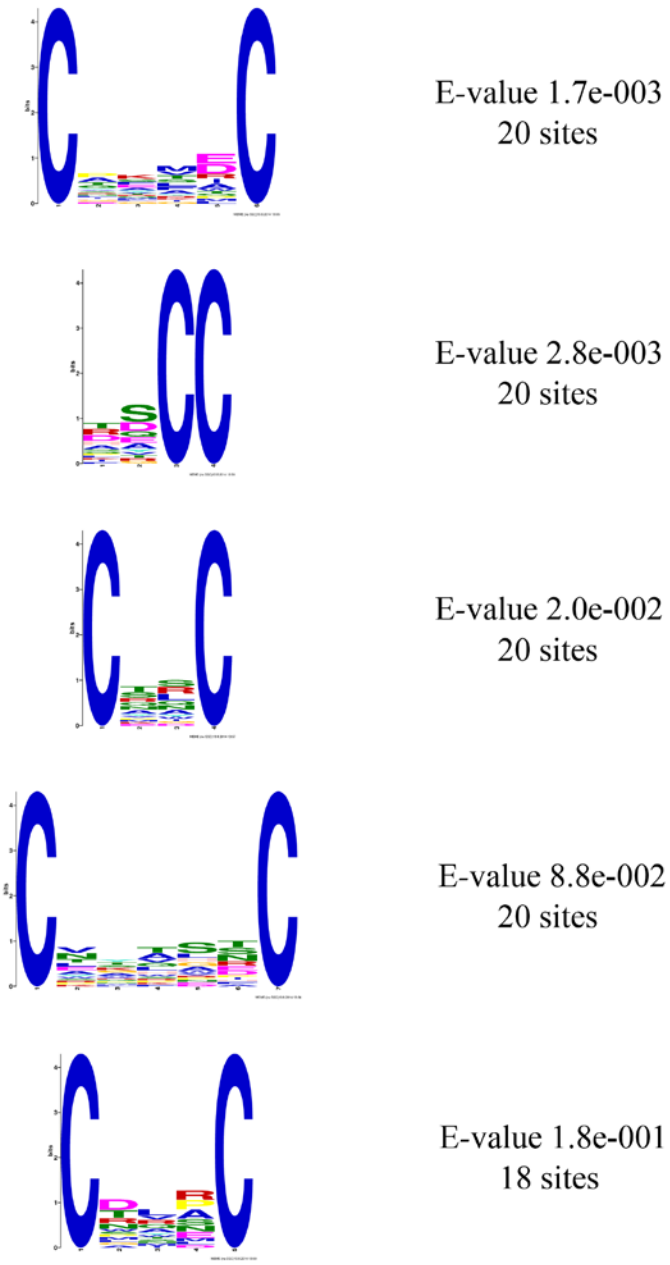


Figure 5-7: ZNF350-binding peptide motifs identified containing two cysteine residues
Five out of the top ten motifs identified through MEME which contain two cysteine residues with the number of sites they are found in and their E-values

5.3 Discussion

Peptide phage display is a powerful tool to efficiently and rapidly identify ligands with affinity for a specific target from a large population of phage clones displaying a diverse range of surface peptides. Large scale identification of binding peptides has been quite restricted due to the limitations of sequencing technologies. However, with the development of next-generation sequencing, it has become possible to sequence thousands of peptides in parallel, enabling the discovery of novel ligands that otherwise may have been lost through multiple rounds of panning in conventional phage display.

Often with peptide phage display, ambiguous false positive results can appear despite apparently successful selection and strong affinity. Binders with no actual affinity for the target are sometimes recovered due to their propagation-based advantages on phage secretion, bacteria infectivity and growth, or binding to other components of the screening system such as the solid phase, capturing reagents or blocking agents rather than the target. The method used in this chapter to analyse and filter specific binding peptides to ZNF350 takes into account the proportion of peptide reads with the target compared to total reads. The phage display carried out in this study had the advantage of multiple different targets analysed and sequenced in parallel, resulting in a vast amount of data available for analysing specificity for ZNF350. By filtering sequences with different parameters including percentage of total reads, fold difference between target and library and/or control the remaining sequences should be specific for the target of interest. To confirm specificity and make sure the sequence of interest isn't a false positive, it is simple to view by eye if the sequence is enriched for any of the other target proteins.

This study used the Ph.D 12-mer library, which is often used when targets may require 7 or fewer defined residues for binding that cannot be contained within the 7-residue window provided by the Ph.D-7 library. This 7-mer library may be more useful for targets that require a short stretch of amino acids for binding and favour few strong contacts as opposed to multiple weak binding contacts. Since ZNF350 is

a transcription factor presumably partaking in multiple transient interactions and binds to at least 7 residues in IRF-1, the 12-mer library was most suitable to ensure the peptides were long enough to allow for ZNF350 binding.

The 12-mer phage display library was screened against ZNF350 to identify peptide motifs that may be the basis for protein-protein interactions. Two distinct motifs were found to show high affinity for ZNF350 with at least a further 5 motifs containing two cysteine residues. Although further experimental validation will be required to confirm binding, by running these motifs through the ScanProsite online program, potentially significant ZNF350 targets have been identified. These include TAF10, an important component of the transcription initiation complex [359] and the histone acetyl transferase p300. p300 binds directly to the C-terminal domain of IRF-1 and leads to the stabilization of p300 binding to p53. This allows for the acetylation of p53, thereby stimulating its activity and leading to the activation of *p21* gene expression. If in fact p300 interacts with ZNF350, it could be that a trimeric complex is formed along with IRF-1, resulting in the acetylation and activation of ZNF350 activity in a similar manner to that displayed with p53. At present, nothing is known about ZNF350 acetylation so it would be of great interest to determine the potential role of p300 binding and its implication on protein function. Indeed, while the relevance of a ZNF350 binding motif being present in these proteins is purely hypothetical, it serves as a basis to explore the possible roles of ZNF350 in a cellular environment.

Chapter 6: Conclusion and future studies

There is very little known about the function of ZNF350, apart from its complex formation with BRCA1 and the subsequent regulation of a handful of genes [43; 45; 52; 54]. When ZNF350 was identified as a possible binding partner of IRF-1 in both a yeast-2 hybrid screen and peptide phage display, it raised the possibility that the two proteins could act together to regulate transcription. IRF-1 plays a role in a multitude of cellular pathways and mostly acts as a positive regulator of expression.

By understanding the relationship between ZNF350 and IRF-1, I hoped to gain insight into their function in relation to gene activity. Most published work on ZNF350 is undertaken with recombinant tagged-ZNF350 so a lot of time was spent attempting to create the optimum conditions for both *in vivo* and *in vitro* analysis of the protein. Once this was established, a variety of techniques were used to successfully demonstrate the formation of the complex within cells and map the ZNF350 binding site to the Mf1 domain of IRF-1, specifically the LXXLL co-regulator motif. The Mf1 domain is located in the C-terminal 25 amino acids and studies have revealed that this region may play a role in the repression of the *Cdk2* gene, the product of which is a component of cell cycle control machinery [99]. Interestingly, it was found that when residues within the motif were mutated, *Cdk2* repression and ultimately growth suppression were lost [98]. It would be of particular interest to identify if this loss of IRF-1 repressor activity may be dependent on ZNF350 binding to the LXXLL motif.

IRF-1 gene expression can be induced by virus, double-stranded RNA, interferons (IFNs) and other cytokines and genotoxic stress [109; 110]. Specifically, IRF-1 protein levels increase upon treatment with IFNγ via activation of the JAK/STAT pathway, or by induction of the ATM signalling pathway by etoposide [109; 112; 113] . Despite the increased levels of IRF-1, its interaction with ZNF350 is reduced, specifically within the nucleus. This would indicate that the formation of the

complex is not induced by cell stress but is instead a feature of the basal state of the cells and may function to maintain homeostasis.

Furthermore, when after treatment with the transcription inhibitor 6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB), or the HDAC inhibitor trichostatin A (TSA), the formation of the ZNF350/IRF-1 complex is reduced, specifically in the nucleus. This however, could be due to the inhibition of *ZNF350* and *IRF-1* gene expression, resulting in lower protein levels. More information on the localisation of the individual proteins within the cells and the conditions they occur will be helpful in further understanding protein function. Furthermore, by identifying genes that are bound and regulated by ZNF350 and IRF-1, it would assist in understanding the role they play in cell processes and perhaps in maintain basal state levels.

By analysing the literature, two genes repressed in a ZNF350-dependent manner were found to contain an ISRE, an IRF-1 recognition site [56; 59; 148; 245]. Matrix metallopeptidase 9 (*MMP9)* is overexpressed in many human malignancies including solid tumours and haematological neoplasms [216-218] and has been identified as a downregulated gene target of ZNF350 [59]. In a separate study, IRF-1 was shown to decrease *MMP9* expression in response to IFNγ by competing with the MMP9-activating NFκB for binding to the promoter [148]. Initial analysis of MMP9 gene activity revealed that ZNF350 and IRF-1 cooperate to repress expression, revealing a novel function of the complex. Since *MMP9* is overexpressed in many human malignancies [216-218] it would be of particular interest to further characterise the role of ZNF350 and IRF-1 in *MMP9* gene repression and under what circumstances this could occur.

The other gene to contain both ZNF350 and IRF-1 putative binding sites is the long terminal repeat of HIV-1. While ZNF350 has been shown to repress HIV-1 LTR activity [56], IRF-1 acts with NFκB to enhance expression [245]. However, after extensive analysis, I determined that both ZNF350 and IRF-1 in fact repress HIV-1 LTR activity and IRF-1 acts independently of ZNF350. Multiple experiments

confirmed this observation so perhaps by analysing the effect of the two proteins on LTR activity in a different manner would be beneficial. It may also be worthwhile to determine the relationship between the role of ZNF350 and IRF-1 with NFκB-mediated HIV-1 LTR activity.

As stated previously, there is limited knowledge of the gene targets of ZNF350 and IRF-1 both individually and cooperatively. Genome-wide studies of protein-DNA interactions and epigenetic marks are essential for a full understanding of transcriptional regulation. An accurate map of binding sites for transcription factors, transcriptional machinery and other DNA binding proteins such as repressors and activators is essential for understanding the networks of gene regulation that determine various biological processes [277]. The main tool for investigating these mechanisms is chromatin immunoprecipitation (ChIP) coupled with either hybridization to a microarray (ChIP-chip) or high-throughput sequencing (ChIP-seq). Following extensive optimization of the methodology of ChIP-seq specific for ZNF350 and IRF-1, the data generated was of too poor quality to gain any significant results from. The actual sequencing of the chromatin was performed adequately, however the immunoprecipitation part of the protocol was somehow compromised. This was very disappointing due to the development of the protocol and the reasons for the problems are still to be determined. Perhaps performing the ChIP-seq in another cell type, for example the commonly used K562 cells and the synthesis of a good-quality ZNF350 monoclonal antibody could improve the results. By tentatively examining the data, three genes were found to bind both ZNF350 and IRF-1 and interestingly these were *ZNF350* and *IRF-1* themselves, along with *MDM2*. The identification of these three genes may be coincidental; however it does appear strange that both *ZNF350* and *IRF-1* genes have been identified. Recently, it was found that the DNA binding domain of IRF-1 is ubiquitinated by MDM2, which acts by binding to the Mf2 domain. Furthermore, IRF-1 is only available for MDM2 binding and subsequent ubiquitination when in its DNA-unbound state, suggesting that IRF-1 degradation may be partly controlled by its ability to bind DNA [313]. ZNF350 is also degraded via the ubiquitin-proteasome pathway [68] and at present the mechanism for this is unknown, however further studies should be carried out to

determine if perhaps MDM2 is involved in this. Providing the ChIP-seq findings can be validated, it would be of interest to examine the relationship between these proteins and their genes further as it may relate to the possible homeostatic activities of the complex and suggest the involvement of a negative feedback loop.

In order to identify gene targets of the complex by another method, a novel screen was developed which involved analysing bioinformatics data on predicted IRF-1 binding regions within the genome and identifying which were in close proximity to a ZNF350 consensus motif. This method successfully identified and validated 16 target genes, 4 of which are involved in the immune response. Specifically, IL-12A was identified as a gene target and previous studies have shown that a loss of IRF-1 has a detrimental effect on LPS-induced IL-12 [266]. Furthermore, IL-12 has been shown to induce IRF-1 gene expression in human NK and T cells [330], which again suggests a possible mechanism of feedback regulation. Follow up studies on the identified targets would help elucidate if indeed the ZNF350/IRF-1 complex acts to maintain homeostasis of the cell or if they could be involved in regulation in response to other factors. The successful identification of potential transcription factor gene targets followed by experimental validation generates a valuable tool for genome-wide studies of gene regulation and can hopefully be developed to provide even more extensive data.

The development of a peptide phage display screen in combination with high-throughput sequencing can preserve the diversity of phage clones and therefore provided a powerful tool for the identification of ZNF350-binding peptides. The peptide lists generated in this study will be further analysed in binding assays such as ELISAs, or alternatively for a more extensive view, by peptide microarrays. Microarrays can immobilise thousands of peptides to enable the characterization of peptide binding hits. However, due to the scale of the phage-display data generated by high-throughput sequencing, this may prove to be impractical. For this reason, the analysis of the data for common motifs was employed in this study to allow for a more directed and manageable study using peptide binding assays.

In conclusion, the work presented here starts to dissect the relationship between ZNF350 and IRF-1 and how this may function within the cellular environment. However, many questions on the occurrence and circumstances of complex formation remain to be investigated in future studies.

Bibliography

1. Hampsey, M. 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* **62**: 465-503.

2. Gaston, K., and Jayaraman, P.S. 2002. Transcriptional repression in eukaryotes: repressors and repression mechanisms. *Cell Mol Life Sci* **60**: 721-741.

3. Baumann, M., Pontiller, J., and Ernst, W. 2010. Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview. *Mol Biotechnol* **45**: 241-247.

4. Kuras, L., and Struhl, K. 1999. Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. *Nature* **399**: 609-613.

5. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. 2002. *Molecular Biology of the Cell*. Garland Science.

6. Sims, R.J., Belotserkovskaya, R., and Reinberg, D. 2004. Elongation by RNA polymerase II: the short and long of it. *Genes & Dev* **18**: 2437-2468.

7. Studitsky, V.M., Walter, W., Kireeva, M., Kashlev, M., Felsenfeld, G. 2004. Chromatin remodeling by RNA polymerases. *Trends in Biomedical Sciences* **29**(3): 127-135.

8. Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., Reinberg, D. 2003. FACT facilitates transcription-dependent nucleosome alteration. *Science* **301**: 1090-1093.

9. Formosa, T., Ruone, S., Adams, M.D., Olsen, A.E., Eriksson, P., Yu, Y., Rhoades, A.R., Kaufman, P.D., and Stillman, D.J. . 2002. Defects in SPT16 or POB3 (yFACT) in Saccharomyces cerevisiae cause dependence on the Hir/Hpc pathway: Polymerase passage may degrade chromatin structure. *Genetics* **162**(4): 1557-1571.

10. Richard, P., and Manley, J.L. 2009. Transcription termination by nuclear RNA polymerases. *Genes & Dev* **23**: 1247-1269.

11. Courey, A.J., and Songtao, J. 2001. Transcriptional repression: the long and short of it. *Genes & Dev* **15**: 2786-2796.

12. Maston, G.A., Evans, S.K., and Green, M.R. 2006. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* **7**: 29-59.

13. Bulger, M., and Groudine, M. 1999. Looping versus linking: toward a model for long-distance gene activation. *Genes & Dev* **13**: 2465-2477.

14. Fuchs, S.Y., Adler, V., Buschmann, T., Wu, X. and Ronai, Z. 1998. Mdm2 association with p53 targets its ubiquitination. *Oncogene* **17**: 2543-2547.

15. Inoue, T., Geyer, R. K., Howard, D., Yu, Z. K. and Maki, C. G. 2001. MDM2 can promote the ubiquitination, nuclear export, and degradation of p53 in the absence of direct binding. *J Biol Chem* **48**: 45255-45260.

16. Hengartner, C.J., Myer, V. E., Liao, S. M., Wilson, C. J., Koh, S. S. and Young, R.A. 1998. Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol Cell* **2**: 43-53.

17. Roth, S.Y., Denu, J.M. and Allis, C.D. 2001. Histone acetyltransferases. *Annu Rev Biochem* **70**: 81-120.

18. Zhang, Y., and Reinberg, D. 2001. Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails. *Genes & Dev* **15**: 2343-2360.

19. Cheung, P., Allis, C.D. and Sassone-Corsi, P. 2000. Signaling to chromatin through histone modifications. *Cell* **103**: 263-271.

20. Jason, L.J., Moore, S.C., Lewis, J.D., Lindsey, G. and Ausio, J. 2002. Histone ubiquitination: A tagging tail unfolds? *BioEssays* **24**: 166-174.

21. Zhang, Y. 2003. Transcriptional regulation by histone ubiquitination and deubiquitination. *Genes & Dev* **17**: 2733-2740.

22. Shiio, Y.a.E., R.N. 2003. Histone sumoylation is associated with transcriptional repression. *PNAS* **100**: 13225–13230.

23. Lewin, B. 2008. *Genes IX*. Jones and Bartlett Publishers.

24. Laity, J.H., ,Lee, B. M. and Wright, P. E. 2001. Zinc finger proteins: new insights into structural amd functional diversity. *Curr Opin Struct Biol* **11**: 39-46.

25. Rosenberg, U.B., Schröder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I., and Jäckle, H. 1986. Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIA. *Nature* **319**: 336-339.

26. Bellefroid, E.J., Poncelet, D. A., Lecocq, P. J., Revelant, O. and Martial, J. A. 1991. The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc Natl Acad Sci U S A* **88**: 3608-3612.

27. Lupo, A., Cesaro, E., Montano, G., Zurlo, D., Izzo, P. and Costanzo, P. 2013. KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions. *Curr Genomics* **14**: 268-278.

28. Urrutia, R. 2003. KRAB-containing zinc-finger repressor proteins. *Genome Biol* **4**(10): 231.

29. Looman, C., Abrink, M., Mark, C. and Hellman, L. 2002. KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. *Mol Biol Evol* **19**: 2118-2130.

30. Wolfe, S.A., Nekludova, L. and Pabo, C. O. 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* **29**: 183-212.

31. Emerson, R.O., and Thomas, J.H. 2009. Adaptive evolution in zinc finger transcription factors. *PLoS Genet* **5**(1): e1000325.

32. Honer, C., Chen, P., Toth, M. J. and Schumacher, C. 2001. Identification of SCAN dimerization domains in four gene families. *Biochim Biophys Acta* **1517**: 441-448.

33. Collins, T., Stone, J. R. and Williams, A. J. 2001. All in the Family: the BTB/POZ, KRAB, and SCAN Domains. *Mol Cell Biol* **21**: 3609-3615.

34. Splettstoesser, T. 2012. Cartoon represntation of C2H2 Zinc finger from X-ray structure of PDB 1A1L. In (ed. http://www.scistyle.com/).

35. Business, S.P. 2007. Representation of the Krüppel Associated Box domain. In (ed. http://commons.wikimedia.org/wiki/File:KRAB_Domain.png).

36. Friedman, J.R., Fredericks, W. J., Jensen, D. E., Speicher, D. W., Huang, X. P., Neilson, E. G. and Rauscher, F. J. III. 1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes & Dev* **10**(16): 2067-2078.

37. Lechner, M.S., Begg, G. E., Speicher, D. W. and Rauscher, F. J. III. 2000. Molecular determinants for targeting heterochromatin protein 1-mediated gene silencing: direct chromoshadow domain-KAP-1 corepressor interaction is essential. . *Mol Cell Biol* **20**: 6449-6465.

38. Ryan, R.F., Schultz, D. C., Ayyanathan, K., Singh, P. B., Friedman, J. R., Fredericks, W. J. and Rauscher, F. J., III. 1999. KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Krüppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. *Mol Cell Biol* **19**(6): 4366-4378.

39. Schultz, D.C., Friedman, J. R. and Rauscher, F. J. III. 2001. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes & Dev* **15**(4): 428-443.

40. Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G. G and Rauscher. F. J. III. 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific

methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Dev* **16**(8): 919-932.

41. Chen, C.F., Li, S., Chen, Y., Chen, P. L., Sharp, Z. D. and Lee, W. H. 1996. The nuclear localization sequences of the BRCA1 protein interact with the importin-alpha subunit of the nuclear transport signal receptor. *J Biol Chem* **271**(50): 32863-32868.

42. Peng, J., and Wysocka, J. 2008. It takes a PHD to SUMO. *Trends Biochem Sci* **33**(5): 191-194.

43. Zheng, L., Pan, H., Li, S., Flesken-Nikitin, A., Chen, P. L., Boyer, T. G., and Lee, W. H. 2000. Sequence-specific transcriptional corepressor function for BRCA1 through a novel zinc finger protein, ZBRK1. *Mol Cell* **6**(4): 757-768.

44. Ran, Q., Wadhwa, R., Bischof, O., Venable, S., Smith, J. R. and Pereira-Smith, O. M. 2001. Characterization of a novel zinc finger gene with increased expression in nondividing normal human cells. *Exp Cell Res* **263**: 156-162.

45. Tan, W., Zheng, L., Lee, W. H., and Boyer, T. G. 2004. Functional dissection of transcription factor ZBRK1 reveals zinc fingers with dual roles in DNA-binding and BRCA1-dependent transcriptional repression. *J Biol Chem* **279**(8): 6576-6587.

46. Tan, W., Kim, S. and Boyer, T. G. 2004. Tetrameric oligomerization mediates transcriptional repression by the BRCA1-dependent Kruppel-associated box-zinc finger protein ZBRK1. *J Biol Chem* **279**(53): 55153-55160.

47. Duncan, J.A., Reeves, J.R., and Cooke, T.G. 1998. BRCA1 and BRCA2 proteins: roles in health and disease. *Molecular pathology* **51**(5): 237-247.

48. Greenberg, R.A., Sobhian, B., Pathania, S., Cantor, S.B., Nakatani, Y., and Livingston, D.M. 2006. Multifactorial contributions to an acute DNA damage response by BRCA1/BARD1-containing complexes. *Genes Dev* **20**(1): 34-46.

49. Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y., Ogata, H., and Ohta, T. . 2001. The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. . *J Biol Chem* **276**: 14537-14540.

50. Wang, X.W., Zhan, Q., Coursen, J. D., Khan, M. A., Kontny, H. U., Yu, L., Hollander, M. C., O'Connor, P. M., Fornace, A. J. Jr and Harris, C. C. 1999. GADD45 induction of a G2/M cell cycle checkpoint. *Proc Natl Acad Sci U S A* **96**(7): 3706-3711.

51. Calderon, M.R., Verway, M., Benslama, R. O., Birlea, M., Bouttier, M., Dimitrov, V., Mader, S.and White, J. H. 2014. Ligand-dependent corepressor contributes to transcriptional repression by C2H2 zinc-finger transcription factor ZBRK1 through association with KRAB-associated protein-1. *Nucleic Acids Res* **gku413**(Epub ahead of print).

52. Furuta, S., Wang, J. M., Wei, S., Jeng, Y. M., Jiang, X., Gu, B., Chen, P. L., Lee, E. Y. and Lee, W. H. 2006. Removal of BRCA1/CtIP/ZBRK1 repressor complex on ANG1 promoter leads to accelerated mammary tumor growth contributed by prominent vasculature. *Cancer Cell* **10**(1): 13-24.

53. Sgarra, R., Rustighi, A., Tessari, M. A., Di Bernardo, J., Altamura, S., Fusco, A., Manfioletti, G. and Giancotti, V. 2004. Nuclear phosphoproteins HMGA and their relationship with chromatin structure and cancer. *FEBS Lett* **574**: 1-8.

54. Ahmed, K.M., Tsai, C. Y. and Lee, W. H. 2010. Derepression of HMGA2 via removal of ZBRK1/BRCA1/CtIP complex enhances mammary tumorigenesis. *J Biol Chem* **285**(7): 4464-4471.

55. Lee, Y.K., Thomas, S. N., Yang, A. J. and Ann, D. K. 2007. Doxorubicin down-regulates Kruppel-associated box domain-associated protein 1 sumoylation that relieves its transcription repression on p21WAF1/CIP1 in breast cancer MCF-7 cells. *J Biol Chem* **282**(3): 1595-1606.

56. Nishitsuji, H., Abe, M., Sawada, R. and Takaku, H. 2012. ZBRK1 represses HIV-1 LTR-mediated transcription. *FEBS Lett* **586**: 3562-3568.

57. Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D. B. and Trono, D. 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**: 237-240.

58. Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M. C. and Shinkai, Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**: 927-931.

59. Lin, L.F., Chuang, C. H., Li, C. F., Liao, C. C., Cheng, C. P., Cheng, T. L., Shen, M. R., Tseng, J. T., Chang, W. C., Lee, W. H and Wang, J. M. 2010. ZBRK1 acts as a metastatic suppressor by directly regulating MMP9 in cervical cancer. *Cancer Res* **70**(1): 192-201.

60. Lin, L.F., Li, C. F., Wang, W. J., Yang, W. M., Wang, D. D., Chang, W. C, Lee, W. H. and Wang, J. M. 2013. Loss of ZBRK1 contributes to the increase of KAP1 and promotes KAP1-mediated metastasis and invasion in cervical cancer. *PLoS One* **8**(8): e73033.

61. Hallen, L., Klein, H., Stoschek, C., Wehrmeyer, S., Nonhoff, U., Ralser, M., Wilde, J., Röhr, C., Schweiger, M. R., Zatloukal, K., Vingron, M., Lehrach, H., Konthur, Z. and Krobitsch, S. 2001. The KRAB-containing zinc-finger transcriptional regulator ZBRK1 activates SCA2 gene transcription through direct interaction with its gene product, ataxin-2. *Hum Mol Genet* **10**(1): 104-114.

62. Yi, Z., Li, Y., Ma, W., Li, D., Zhu, C., Luo, J., Wang, Y., Huang, X., Yuan, W., Liu, M. and Wu, X. 2004. A novel KRAB zinc-finger protein, ZNF480, expresses in human heart and activates transcriptional activities of AP-1 and SRE. *Biochem Biophys Res Commun* **320**(2): 409-415.

63. Garcia, V., Domínguez, G., García, J. M., Silva, J., Peña, C., Silva, J. M., Carcereny, E., Menendez, J., España, P. and Bonilla, F. 2004. Altered expression of the ZBRK1 gene in human breast carcinomas. *J Pathol* **202**(2): 224-232.

64. Garcia, V., García, J. M., Peña, C., Silva, J., Domínguez, G., Rodríguez, R., Maximiano, C., Espinosa, R., España, P.and Bonilla, F. 2005. The GADD45, ZBRK1 and BRCA1 pathway: quantitative analysis of mRNA expression in colon carcinomas. *J Pathol* **206**(1): 92-99.

65. Ensembl. 2014. Gene: ZNF350. In.

66. Plourde, K.V., Labrie, Y., Desjardins, S., Belleau, P., Ouellette, G., Durocher, F. and INHERIT BRCAs. 2013. Analysis of ZNF350/ZBRK1 promoter variants and breast cancer susceptibility in non-BRCA1/2 French Canadian breast cancer families. *J Hum Genet* **58**(2): 59-66.

67. Desjardins, S., Belleau, P., Labrie, Y., Ouellette, G., Bessette, P., Chiquette, J., Laframboise, R., Lépine, J., Lespérance, B., Pichette, R., Plante, M; INHERIT BRCAs and Durocher, F. 2008. Genetic variants and haplotype analyses of the ZBRK1/ZNF350 gene in high-risk non BRCA1/2 French Canadian breast and ovarian cancer families. *Int J Cancer* **122**(1): 108-116.

68. Yun, J., and Lee, W. H. 2003. Degradation of transcription repressor ZBRK1 through the ubiquitin-proteasome pathway relieves repression of Gadd45a upon DNA damage. *Mol Cell Biol* **23**(20): 7305-7314.

69. Yun, M.H., and Hiom, K. 2009. Understanding the functions of BRCA1 in the DNA-damage response. *Biochem Soc Trans* **37**: 597-604.

70. Chellappan, S.P., Hiebert, S., Mudryj, M., Horowitz, J. M. and Nevins, J. R. 1991. The E2F transcription factor is a cellular target for the RB protein. *Cell* **65**(6): 1053-1061.

71. Liao, C.C., Tsai, C. Y., Chang, W. C., Lee, W. H. and Wang, J. M. 2010. RB·E2F1 complex mediates DNA damage responses through transcriptional regulation of ZBRK1. *J Biol Chem* **285**(43): 33134-33143.

72. Taniguchi, T., Ogasawara, K., Takaoka, A. and Tanaka, N. 2001. IRF family of transcription factors as regulators of host defense. *Annu Rev Immunol* **19**: 623-655.

73. Tamura, T., Yanai, H., Savitsky, D., and Taniguchi, T. 2008. The IRF family transcription factors in immunity and oncogenesis. *Annu Rev Immunol* **26**: 535-584.

74. Fujii, Y., Shimizu, T., Kusumoto, M., Kyogoku, Y., Taniguchi, T., and Hakoshima, T. 1999. Crystal structure of an IRF-DNA complex reveals novel DNA recognition and cooperative binding to a tandem repeat of core sequences. . *EMBO J* **18**(18): 5028-5041.

75. Nguyen, H., Hiscott, J., and Pitha, P.M. 1997. The growing family of interferon regulatory factors. *Cytokine Growth Factor Rev* **8**(4): 293-312.

76. Narayan, V., Eckert, M., Zylicz, A., Zylicz, M., and Ball, K.L. 2009. Cooperative regulation of the interferon regulatory factor-1 tumor suppressor protein by core components of the molecular chaperone machinery. *J Biol Chem* **284**(38): 25889-25899.

77. Tamura, T., Ishihara, M., Lamphier, M.S., Tanaka, N., Oishi, I., Aizawa, S., Matsuyama, T., Mak, T.W., Taki, S., and Taniguchi, T. 1997. DNA damage-induced apoptosis and Ice gene induction in mitogenically activated T lymphocytes require IRF-1. *Leukemia* **11**: 439-440.

78. Harada, H., Fujita, T., Miyamoto, M., Kimura, Y., Maruyama, M., Furia, A., Miyata, T., and Taniguchi, T. 1989. Structurally similar but functionally distinct factors, IRF-1 and IRF-2, bind to the same regulatory elements of IFN and IFN-inducible genes. *Cell* **58**(4): 729-739.

79. Nozawa, H., Oda, E., Nakao, K., Ishihara, M., Ueda, S., Yokochi, T., Ogasawara, K., Nakatsuru, Y., Shimizu, S., Ohira, Y., Hioki, K., Aizawa, S., Ishikawa, T., Katsuki, M., Muto, T., Taniguchi, T., and Tanaka, N. 1999. Loss of transcription factor IRF-1 affects tumor susceptibility in mice carrying the Ha-ras transgene or nullizygosity for p53. *Genes & Dev* **13**(10): 1240-1245.

80. Yanai, H., Negishi, H., and Taniguchi, T. 2012. The IRF family of transcription factors: Inception, impact and implications in oncogenesis. . *Oncoimmunology* **1**(8): 1376-1386.

81. Lohoff, M., and Mak, T.W. 2005. Roles of interferon-regulatory factors in T-helper-cell differentiation. *Nat Rev Immunol* **5**(2): 125-135.

82. Schaper, F., Kirchhoff, S., Posern, G., Koster, M., Oumard, A., Sharf, R., Levi, B.Z., Hauser, H. 1998. Functional domains of interferon regulatory factor I (IRF-1). *Biochem J* **335**: 147-157.

83. Eason, D.D., Shepherd, A.T., and Blanck, G. 1999. Interferon regulatory factor 1 tryptophan 11 to arginine point mutation abolishes DNA binding. *Biochim Biophys Acta* **1446**(1-2): 140-144.

84. Dornan, D., Eckert, M., Wallace, M., Shimizu, H., Ramsay, E., Hupp, T. R., and Ball, K. L. 2004. Interferon regulatory factor 1 binding to p300 stimulates DNA-dependent acetylation of p53. *Mol Cell Biol* **24**(22): 10083-10098.

85. Escalante, C.R., Yie, J., Thanos, D., and Aggarwal, A.K. 1998. Structure of IRF-1 with bound DNA reveals determinants of interferon regulation. *Nature* **391**: 103-106.

86. Spink, J., and Evans, T. 1997. Binding of the transcription factor interferon regulatory factor-1 to the inducible nitric-oxide synthase promoter. *J Biol Chem* **272**(39): 24417-24425.

87. Narayan, V., Halada, P., Hernychova, L., Chong, Y.P., Zakova, J., Hupp, T.R., Vojtesek, B., and Ball, K.L. . 2011. A multiprotein binding interface in an intrinsically disordered region of the tumor suppressor protein interferon regulatory factor-1. *J Biol Chem* **286**(16): 14291-14303.

88. Narayan, V., Pion, E., Landre, V., Muller, P., and Ball, K.L. 2011. Docking-dependent ubiquitination of the interferon regulatory factor-1 tumor suppressor protein by the ubiquitin ligase CHIP. *J Biol Chem* **286**(1): 607-619.

89. Landré, V., Pion, E., Narayan, V., Xirodimas, D.P., and Ball, K.L. 2013. DNA-binding regulates site-specific ubiquitination of IRF-1. *Biochem J* **449**(3): 707-717.

90. Lin, R., Mustafa, A., Nguyen, H., Gewert, D. and Hiscott, J. 1994. Mutational analysis of interferon (IFN) regulatory factors 1 and 2. Effects on the induction of IFN-beta gene expression. *J Biol Chem* **269**(26): 17542-17549.

91. Moore, P.A., Ruben, S. M. and Rosen, C. A. 1993. Conservation of transcriptional activation functions of the NF-kappa B p50 and p65 subunits in mammalian cells and Saccharomyces cerevisiae. *Mol Cell Biol* **13**(3): 1666-1674.

92. Ma, J., and Ptashne, M. 1987. A new class of yeast transcriptional activators. *Cell* **51**(1): 113-119.

93. Kirchhoff, S., Oumard, A., Nourbakhsh, M., Levi, B.Z., and Hauser, H. 2000. Interplay between repressing and activating domains defines the transcriptional activity of IRF-1. *Eur J Biochem* **267**: 6753-6761.

94. Kim, E.J., Park, C.H., Park, J.S., and Um, S.J. 2003. Functional dissection of the transactivation domain of interferon regulatory factor-1. *Biochem Biophys Res Commun* **304**: 253-259.

95. Weisz, A., Marx, P., Sharf, R., Appella, E., Driggers, P.H., Ozato, K., and Levi, B.Z. 1992. Human interferon consensus sequence binding protein is a negative regulator of enhancer elements common to interferon-inducible genes. *J Biol Chem* **267**(35): 25589-25596.

96. Nelson, N., Marks, M.S., Driggers, P.H., and Ozato, K. 1993. Interferon consensus sequence-binding protein, a member of the interferon regulatory factor family, suppresses interferon-induced gene transcription. *Mol Cell Biol* **13**(1): 588-599.

97. Kirchhoff, S., Schaper, F., Oumard, A., and Hauser, H. 1998. In vivo formation of IRF-1 homodimers. *Biochimie* **80**: 659-664.

98. Eckert, M., Meek, S.E., and Ball, K.L. 2006. A novel repressor domain is required for maximal growth inhibition by the IRF-1 tumor suppressor. *J Biol Chem* **281**(32): 23092-23102.

99. Xie, R.L., Gupta, S., Miele, A., Shiffman, D., Stein, J.L., Stein, G.S., and van Wijnen, A.J. 2003. The tumor suppressor interferon regulatory factor 1 interferes with SP1 activation to repress the human CDK2 promoter. *J Biol Chem* **278**(29): 26589-26596.

100. Plevin, M.J., Mills, M.M., and Ikura, M. 2005. The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem Sci* **30**(2): 66-69.

101. Pion, E., Narayan, V., Eckert, M., and Ball, K.L. 2009. Role of the IRF-1 enhancer domain in signalling polyubiquitination and degradation. *Cell Signal* **21**(10): 1479-1487.

102. Moller, A., Pion, E., Narayan, V., and Ball, K.L. 2010. Intracellular activation of interferon regulatory factor-1 by nanobodies to the multifunctional (Mf1) domain. *J Biol Chem* **285**(49): 38348-38361.

103. Thomas, D., and Tyers, M. . 2000. Transcriptional regulation: Kamikaze activators. *Curr Biol* **10**(9): R341-343.

104. Willman, C.L., Sever, C.E., Pallavicini, M.G., Harada, H., Tanaka, N., Slovak, M.L., Yamamoto, H., Harada, K., Meeker, T.C., List, A.F., et al. 1993. Deletion of IRF-1, mapping to chromosome 5q31.1, in human leukemia and preleukemic myelodysplasia. *Science* **259**(5097): 968-971.

105. Sims, S.H., Cha, Y., Romine, M.F., Gao, P.Q., Gottlieb, K., and Deisseroth, A.B. 1993. A novel interferon-inducible domain: structural and functional analysis of the human interferon regulatory factor 1 gene promoter. *Mol Cell Biol* **13**(1): 690-702.

106. Harada, H., Takahashi, E., Itoh, S., Harada, K., Hori, T.A., and Taniguchi, T. 1994. Structure and regulation of the human interferon regulatory factor 1 (IRF-1) and IRF-2 genes: implications for a gene network in the interferon system. *Mol Cell Biol* **14**(2): 1500-1509.

107. Lenardo, M.J., Fan, C.M., Maniatis, T., and Baltimore, D. 1989. The involvement of NF-kappa B in beta-interferon gene regulation reveals its role as widely inducible mediator of signal transduction. *Cell* **57**(2): 287-294.

108. Visvanathan, K.V., and Goodbourn, S. 1989. Double-stranded RNA activates binding of NF-kappa B to an inducible element in the human beta-interferon promoter. *EMBO J* **8**(4): 1129-1138.

109. Pamment, J., Ramsay, E., Kelleher, M., Dornan, D., and Ball, K.L. 2002. Regulation of the IRF-1 tumour modifier during the response to genotoxic stress involves an ATM-dependent signalling pathway. *Oncogene* **21**(51): 7776-7785.

110. Romeo, G., Fiorucci, G., Chiantore, M.V., Percario, Z.A., Vannucchi, S., and Affabris, E. 2002. IRF-1 as a negative regulator of cell proliferation. *J Interferon Cytokine Res* **22**(1): 39-47.

111. Taniguchi, T., Tanaka, N., and Taki, S. 1998. Regulation of the interferon system, immune response and oncogenesis by the transcription factor interferon regulatory factor-1. *Eur Cytokine Netw* **9**: 43-48.

112. Decker, T., Lew, D.J., and Darnell, J.E., Jr. 1991. Two distinct alpha-interferon-dependent signal transduction pathways may contribute to activation of transcription of the guanylate-binding protein gene. *Mol Cell Biol* **11**(10): 5147-5153.

113. Takaoka, A.a.Y., H. 2006. Interferon signalling network in innate defence. *Cell Microbiol* **8**(6): 907-922.

114. Tanaka, N., Ishihara, M., Lamphier, M.S., Nozawa, H., Matsuyama, T., Mak, T.W., Aizawa, S., Tokino, T., Oren, M., and Taniguchi, T. 1996. Cooperation of the

tumour suppressors IRF-1 and p53 in response to DNA damage. *Nature* **382**(6594): 816-818.

115. Janeway, C.A.J., and Medzhitov, R. 2002. Innate immune recognition. *Annu Rev Immunol* **20**: 197-216.

116. Kirchhoff, S., Koromilas, A.E., Schaper, F., Grashoff, M., Sonenberg, N., and Hauser, H. 1995. IRF-1 induced cell growth inhibition and interferon induction requires the activity of the protein kinase PKR. *Oncogene* **11**(3): 439-445.

117. Su, Z.Z., Sarkar, D., Emdad, L., Barral, P.M., and Fisher, P.B. 2007. Central role of interferon regulatory factor-1 (IRF-1) in controlling retinoic acid inducible gene-I (RIG-I) expression. *J Cell Physiol* **213**(2): 502-510.

118. Negishi, H., Fujita, Y., Yanai, H., Sakaguchi, S., Ouyang, X., Shinohara, M., Takayanagi, H,, Ohba, Y., Taniguchi, T., and Honda, K. 2006. Evidence for licensing of IFN-gamma-induced IFN regulatory factor 1 transcription factor by MyD88 in Toll-like receptor-dependent gene induction program. *Proc Natl Acad Sci U S A* **103**(41): 15136-15141.

119. Kamijo, R., Harada, H., Matsuyama, T., Bosland, M., Gerecitano, J., Shapiro, D., Le, J., Koh, S.I., Kimura, T., Green, S.J., et al. 1994. Requirement for transcription factor IRF-1 in NO synthase induction in macrophages. *Science* **263**(5153): 1612-1615.

120. Tamura, T., Ishihara, M., Lamphier, M.S., Tanaka, N., Oishi, I., Aizawa, S., Matsuyama, T., Mak, T.W., Taki, S., and Taniguchi, T. 1995. An IRF-1-dependent pathway of DNA damage-induced apoptosis in mitogen-activated T lymphocytes. *Nature* **376**(6541): 596-599.

121. Martin, E., Nathan, C., and Xie, Q.W. 1994. Role of interferon regulatory factor 1 in induction of nitric oxide synthase. *J Exp Med* **180**(3): 977-984.

122. Tada, Y., Ho, A., Matsuyama, T., and Mak, T.W. 1997. Reduced incidence and severity of antigen-induced autoimmune diseases in mice lacking interferon regulatory factor-1. *J Exp Med* **185**(2): 231-238.

123. Nguyen, H., Teskey, L., Lin, R., and Hiscott, J. 1999. Identification of the secretory leukocyte protease inhibitor (SLPI) as a target of IRF-1 regulation. *Oncogene* **18**(39): 5455-5463.

124. Blanco, J.C., Contursi, C., Salkowski, C.A., DeWitt, D.L., Ozato, K., and Vogel, S.N. 2000. Interferon regulatory factor (IRF)-1 and IRF-2 regulate interferon gamma-dependent cyclooxygenase 2 expression. *J Exp Med* **191**(2): 2131-2144.

125. Neish, A.S., Read, M.A., Thanos, D., Pine, R., Maniatis, T., and Collins, T. 1995. Endothelial interferon regulatory factor 1 cooperates with NF-kappa B as a transcriptional activator of vascular cell adhesion molecule 1. *Mol Cell Biol* **15**(5): 2558-2569.

126. Marsili, G., Borsetti, A., Sgarbanti, M., Remoli, A.L., Ridolfi, B., Stellacci, E., Ensoli, B., and Battistini, A. 2003. On the role of interferon regulatory factors in HIV-1 replication. *Ann N Y Acad Sci* **1010**: 29-42.

127. Fowke, K.R., Nagelkerke, N.J., Kimani, J., Simonsen, J.N., Anzala, A.O., Bwayo, J.J., MacDonald, K.S., Ngugi, E.N., and Plummer, F.A. . 1996. Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. *Lancet* **348**: 1347-1351.

128. Plummer, F.A., Ball, T.B., Kimani, J., and Fowke, K.R. 1999. Resistance to HIV-1 infection among highly exposed sex workers in Nairobi: what mediates protection and why does it develop? *Immunol Lett* **66**: 27-34.

129. Ball, T.B., Ji, H., Kimani, J., McLaren, P., Marlin, C., Hill, A.V., and Plummer, F.A. 2007. Polymorphisms in IRF-1 associated with resistance to HIV-1 infection in highly exposed uninfected Kenyan sex workers. *AIDS* **21**(9): 1091-1101.

130. Ji, H., Ball, T.B., Ao, Z., Kimani, J., Yao, X., and Plummer, F.A. 2010. Reduced HIV-1 long terminal repeat transcription in subjects with protective interferon regulatory factor-1 genotype: a potential mechanism mediating resistance to infection by HIV-1. *Scand J Infect Dis* **42**(5): 389-394.

131. Boultwood, J., Fidler, C., Lewis, S., MacCarthy, A., Sheridan, H., Kelly, S., Oscier, D., Buckle, V.J., and Wainscoat. J.S. . 1993. Allelic loss of IRF1 in myelodysplasia and acute myeloid leukemia: retention of IRF1 on the 5q-chromosome in some patients with the 5q-syndrome. *Blood* **82**: 2611-2616.

132. Ogasawara, S., Tamura, G., Maesawa, C., Suzuki, Y., Ishida, K., Satoh, N., Uesugi, N., Saito, K., and Satodate, R. 1996. Common deleted region on the long arm of chromosome 5 in esophageal carcinoma. *Gastroenterology* **110**: 52-57.

133. Tamura, G., Ogasawara, S., Nishizuka, S., Sakata, K., Maesawa, C., Suzuki, Y., Terashima, M., Saito, K., and Satodate, R. 1996. Two distinct regions of deletion on the long arm of chromosome 5 in differentiated adenocarcinomas of the stomach. *Cancer Res* **56**(3): 612-615.

134. Nozawa, H., Oda, E., Ueda, S., Tamura, G., Maesawa, C., Muto, T., Taniguchi, T., and Tanaka, N. . 1998. Functionally inactivating point mutation in the tumor-suppressor IRF-1 gene identified in human gastric cancer. *Int J Cancer* **77**(4): 522-527.

135. Connett, J.M., Badri, L., Giordano, T.J., Connett, W.C., and Doherty, G.M. 2005. Interferon regulatory factor 1 (IRF-1) and IRF-2 expression in breast cancer tissue microarrays. *J Interferon Cytokine Res* **25**(10): 587-594.

136. Kondo, T., Minamino, N., Nagamura-Inoue, T., Matsumoto, M., Taniguchi, T., and Tanaka, N. 1997. Identification and characterization of nucleophosmin/B23/numatrin which binds the anti-oncogenic transcription factor IRF-1 and manifests oncogenic activity. *Oncogene* **15**(11): 1275-1281.

137. Park, J.S., Kim, E.J., Kwon, H.J., Hwang, E.S., Namkoong, S.E., and Um, S.J. 2000. Inactivation of interferon regulatory factor-1 tumor suppressor protein by

HPV E7 oncoprotein. Implication for the E7-mediated immune evasion mechanism in cervical carcinogenesis. *J Biol Chem* **275**(10): 6764-6769.

138. Cesarman, E., Chang, Y., Moore, P.S., Said, J.W., and Knowles, D.M. 1995. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. *N Engl J Med* **332**(18): 1186-1191.

139. Chang, Y., Cesarman, E., Pessin, M.S., Lee, F., Culpepper, J., Knowles, D.M., and Moore, P.S. 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**(5192): 1865-1869.

140. Soulier, J., Grollet, L., Oksenhendler, E., Cacoub, P., Cazals-Hatem, D., Babinet, P., d'Agay, M.F., Clauvel, J.P., Raphael, M., Degos, L., et al. 1995. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. *Blood* **86**(4): 1276-1280.

141. Russo, J.J., Bohenzky, R.A., Chien, M.C., Chen, J., Yan, M., Maddalena, D., Parry, J.P., Peruzzi, D., Edelman, I.S., Chang, Y., and Moore, P.S. 1996. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *Proc Natl Acad Sci U S A* **93**(5): 14862-14867.

142. Cunningham, C., Barnard, S., Blackbourn, D.J., and Davison, A.J. 2003. Transcription mapping of human herpesvirus 8 genes encoding viral interferon regulatory factors. *J Gen Virol* **84**(Pt 6): 1471-1483.

143. Li, M., Damania, B., Alvarez, X., Ogryzko, V., Ozato, K., and Jung, J.U. 2000. Inhibition of p300 histone acetyltransferase by viral interferon regulatory factor. *Mol Cell Biol* **20**(21): 8254-8263.

144. Lin, R., Genin, P., Mamane, Y., Sgarbanti, M., Battistini, A., Harrington, W.J. Jr., Barber, G.N., and Hiscott, J. 2001. HHV-8 encoded vIRF-1 represses the interferon antiviral response by blocking IRF-3 recruitment of the CBP/p300 coactivators. *Oncogene* **20**(7): 800-811.

145. Tanaka, N., Ishihara, M., Kitagawa, M., Harada, H., Kimura, T., Matsuyama, T., Lamphier, M.S., Aizawa, S., Mak, T.W., and Taniguchi, T. 1994. Cellular commitment to oncogene-induced transformation or apoptosis is dependent on the transcription factor IRF-1. *Cell* **77**(6): 829-839.

146. Passioura, T., Dolnikov, A., Shen, S., and Symonds, G. 2005. N-ras-induced growth suppression of myeloid cells is mediated by IRF-1. . *Cancer Res* **65**(3): 797-804.

147. Kupferman, M.E., Fini, M.E., Muller, W.J., Weber, R., Cheng, Y., and Muschel, R.J. 2000. Matrix metalloproteinase 9 promoter activity is induced coincident with invasion during tumor progression. *Am J Pathol* **157**(6): 1777-1783.

148. Sancéau, J., Boyd, D.D., Seiki, M., and Bauvois, B. 2002. Interferons inhibit tumor necrosis factor-alpha-mediated matrix metalloproteinase-9 activation via interferon regulatory factor-1 binding competition with NF-kappa B. *J Biol Chem* **277**(38): 35766-35775.

149. Harada, H., Kitagawa, M., Tanaka, N., Yamamoto, H., Harada, K., Ishihara, M., and Taniguchi, T. 1993. Anti-oncogenic and oncogenic potentials of interferon regulatory factors-1 and -2. *Science* **259**(5097): 971-974.

150. Strasser, A., Harris, A.W., Jacks, T., and Cory, S. 1994. DNA damage can induce apoptosis in proliferating lymphoid cells via p53-independent mechanisms inhibitable by Bcl-2. . *Cell* **79**: 329-339.

151. Lallemand, C., Palmieri, M., Blanchard, B., Meritet, J.F., and Tovey, M.G. 2002. GAAP-1: a transcriptional activator of p53 and IRF-1 possesses pro-apoptotic activity. *EMBO Rep* **3**(2): 153-158.

152. Huang, Y., Walstrom, A., Zhang, L., Zhao, Y., Cui, M., Ye, L., and Zheng, J.C. 2009. Type I interferons and interferon regulatory factors regulate TNF-related apoptosis-inducing ligand (TRAIL) in HIV-1-infected macrophages. *PLoS One* **4**(4): e5397.

153. Stang, M.T., Armstrong, M.J., Watson, G.A., Sung, K.Y., Liu, Y., Ren, B., and Yim, J.H. 2007. Interferon regulatory factor-1-induced apoptosis mediated by a ligand-independent fas-associated death domain pathway in breast cancer cells. *Oncogene* **26**(44): 6420-6430.

154. Gao, J., Senthil, M., Ren, B., Yan, J., Xing, Q., Yu, J., Zhang, L., and Yim, J.H. 2010. IRF-1 transcriptionally upregulates PUMA, which mediates the mitochondrial apoptotic pathway in IRF-1-induced apoptosis in cancer cells. *Cell Death Differ* **17**(4): 699-709.

155. Wang, J., Zhang, W., Zhang, Y., Chen, Y., Zou, B., Jiang, B., Pang, R., Gu, Q., Qiao, L., Lan, H., Kung, H.F., and Wong, B.C. 2009. c-Jun N-terminal kinase (JNK1) upregulates XIAP-associated factor 1 (XAF1) through interferon regulatory factor 1 (IRF-1) in gastrointestinal cancer. *Carcinogenesis* **30**(2): 222-229.

156. TheGenePool. 2009. http://genepool.bio.ed.ac.uk/. In.

157. Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J., and Nekrutenko, A. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* **15**(10): 1451-1455.

158. Princeton. 2013. Tutorials: ChIP-Seq. In (ed. https://sites.google.com/site/princetonhtseq/tutorials/chip-seq).

159. BaCR. 2013. Analysis of ChIP-seq data in Galaxy. In (ed. http://jura.wi.mit.edu/bio/education/hot_topics/galaxy/GalaxyNov2012_ChIP-seq_toPost.pdf).

160. Moller, A. 2010. Development of an intrabody capable of activating interferon regulatroy factor-1 (IRF-1) and identification of IRF-1 binding peptide motifs. In *College of Medicine and Veterinary Medicine*. University of Edinburgh, Edinburgh.

161. Narayan, V. 2011. Interactomics identifies novel regulators of the IRF-1 tumour suppressor protein. In *School of Molecular, Genetic and Population Health Sciences*. The University of Edinburgh.

162. Litterst, C.M., and Pfitzner, E. 2002. An LxxLL motif in the transactivation domain of STAT6 mediates recruitment of NCoA-1/SRC-1. *J Biol Chem* **277**: 36052-36060.

163. Torchia, J., Rose, D.W., Inostroza, J., Kamei, Y., Westin, S., Glass, C.K., and Rosenfeld, M.G. 1997. The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. *Nature* **387**(6634): 677-684.

164. Fujita, T., Kimura, Y., Miyamoto, M., Barsoumian, E.L., and Taniguchi, T. 1989. Induction of endogenous IFN-alpha and IFN-beta genes by a regulatory transcription factor, IRF-1. *Nature* **337**(6204): 270-272.

165. Buro, L.J., Chipumuro, E., and Henriksen, M.A. 2010. Menin and RNF20 recruitment is associated with dynamic histone modifications that regulate signal transducer and activator of transcription 1 (STAT1)-activated transcription of the interferon regulatory factor 1 gene (IRF1). *Epigenetics Chromatin* **3**(1): 16.

166. Yankulov, K., Yamashita, K., Roy, R., Egly, J.M., and Bentley, D.L. 1995. The transcriptional elongation inhibitor 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription factor IIH-associated protein kinase. *J Biol Chem* **270**(41): 23922-23925.

167. Marshall, N.F., and Price, D.H. 1995. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem* **270**(21): 12335-12338.

168. Koumenis, C., and Giaccia, A. 1997. Transformed cells require continuous activity of RNA polymerase II to resist oncogene-induced apoptosis. *Mol Cell Biol* **17**(12): 7306-7316.

169. Ljungman, M., Zhang, F., Chen, F., Rainbow, A.J., and McKay, B.C. 1999. Inhibition of RNA polymerase II as a trigger for the p53 response. *Oncogene* **18**(3): 583-592.

170. Dubois, M.F., Nguyen, V.T., Bellier, S., and Bensaude, O. 1994. Inhibitors of transcription such as 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole and isoquinoline sulfonamide derivatives (H-8 and H-7) promote dephosphorylation of the carboxyl-terminal domain of RNA polymerase II largest subunit. *J Biol Chem* **269**(18): 13331-13336.

171. te Poele, R.H., Okorokov, A.L., and Joel, S.P. 1999. RNA synthesis block by 5, 6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) triggers p53-dependent apoptosis in human colon carcinoma cells. *Oncogene* **18**(42): 5765-5772.

172. Ljungman, M., and Zhang, F. 1996. Blockage of RNA polymerase as a possible trigger for u.v. light-induced apoptosis. *Oncogene* **13**(4): 823-831.

173. Blaydes, J.P., Craig, A.L., Wallace, M., Ball, H.M., Traynor, N.J., Gibbs, N.K., and Hupp, T.R. 2000. Synergistic activation of p53-dependent transcription by two cooperating damage recognition pathways. *Oncogene* **19**(34): 3829-3839.

174. Ljungman, M., O'Hagan, H.M., Paulsen, M.T. 2001. Induction of ser15 and lys382 modifications of p53 by blockage of transcription elongation. *Oncogene* **20**(42): 5964-5971.

175. Blattner, C., Sparks, A., and Lane, D. 1999. Transcription factor E2F-1 is upregulated in response to DNA damage in a manner analogous to that of p53. *Mol Cell Biol* **19**(5): 3704-3713.

176. Martin, K., Trouche, D., Hagemeier, C., Sorensen, T.S., LaThangue, N.B., and Kouzarides, T. 1995. Stimulation of E2F-11/DP-1 transcriptional activity by Mdm2 oncoprotein. *Nature* **375**: 691-694.

177. Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G. and Landegren, U.

2006. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nature Methods* **3**: 995-1000.

178. Gullberg, M., and Andersson, A.C. 2010. Visualization and quantification of protein-protein interactions in cells and tissues. *Nature Methods* **7**.

179. Ishii, K.J., Koyama, S., Nakagawa, A., Coban, C., and Akira, S. 2008. Host innate immune receptors and beyond: making sense of microbial infections. *Cell Host Microbe* **3**: 352-363.

180. McCartney, S., Vermi, W., Gilfillan, S., Cella, M., Murphy, T.L., Schreiber, R.D., Murphy, K.M., and Colonna, M. 2009. Distinct and complementary functions of MDA5 and TLR3 in poly(I:C)-mediated activation of mouse NK cells. *J Exp Med* **206**(13): 2967-2976.

181. Pine, R. 1997. Convergence of TNFalpha and IFNgamma signalling pathways through synergistic induction of IRF-1/ISGF-2 is mediated by a composite GAS/kappaB promoter element. *Nucleic Acids Res* **25**(21): 4346-4354.

182. Gupta, S., Xia, D., Jiang, M., Lee, S., and Pernis, A.B. 1998. Signaling pathways mediated by the TNF- and cytokine-receptor families target a common cis-element of the IFN regulatory factor 1 promoter. *J Immunol* **161**(11): 5997-6004.

183. Heitmeier, M.R., Scarim, A.L., and Corbett, J.A. 1998. Double-stranded RNA-induced inducible nitric-oxide synthase expression and interleukin-1 release by murine macrophages requires NF-kappaB activation. *J Biol Chem* **273**(24): 15301-15307.

184. Stark, G.R., Kerr, I.M., Williams, B.R., Silverman, R.H., and Schreiber, R.D. 1998. How cells respond to interferons. *Annu Rev Biochem* **67**: 227-264.

185. Blair, L.A., Maggi, L.B. Jr., Scarim, A.L., and Corbett, J.A. 2002. Role of interferon regulatory factor-1 in double-stranded RNA-induced iNOS expression by mouse islets. *J Biol Chem* **277**(1): 359-365.

186. Pommier, Y., Leo, E., Zhang, H., and Marchand, C. 2010. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol* **17**(5): 421-433.

187. Feng, Z., Kachnic, L., Zhang, J., Powell, S.N., and Xia, F. 2004. DNA damage induces p53-dependent BRCA1 nuclear export. *J Biol Chem* **279**(27): 28574-28584.

188. Jiang, J., Yang, E.S., Jiang, G., Nowsheen, S., Wang, H., Wang, T., Wang, Y., Billheimer, D., Chakravarthy, A.B., Brown, M., Haffty, B., and Xia, F. 2011. p53-dependent BRCA1 nuclear export controls cellular susceptibility to DNA damage. *Cancer Res* **71**(16): 5546-5557.

189. Jin, S., Antinore, M.J., Lung, F.D., Dong, X., Zhao, H., Fan, F., Colchagie, A.B., Blanck, P., Roller, P.P., Fornace, A.J. Jr., and Zhan, Q. 2000. The GADD45 inhibition of Cdc2 kinase correlates with GADD45-mediated growth suppression. *J Biol Chem* **275**(22): 16602-16608.

190. Wang, X.W., Zhan, Q., Coursen, J.D., Khan, M.A., Kontny, H.U., Yu, L., Hollander, M.C., O'Connor, P.M., Fornace, A.J. Jr., and Harris, C.C. 1999. GADD45 induction of a G2/M cell cycle checkpoint. *Proc Natl Acad Sci U S A* **96**(7): 3706-3711.

191. Arriola, E.L., Lopez, A.R., and Chresta, C.M. 1999. Differential regulation of p21waf-1/cip-1 and Mdm2 by etoposide: etoposide inhibits the p53-Mdm2 autoregulatory feedback loop. *Oncogene* **18**(4): 1081-1091.

192. Li, S., Chen, P. L., Subramanian, T., Chinnadurai, G., Tomlinson, G., Osborne, C. K., Sharp, Z. D. and Lee, W. H. 1999. Binding of CtIP to the BRCT repeats of BRCA1 involved in the transcription regulation of p21 is disrupted upon DNA damage. *J Biol Chem* **274**(16): 11334-11338.

193. Turinetto, V., Porcedda, P., Orlando, L., De Marchi, M., Amoroso, A., and Giachino, C. 2009. The cyclin-dependent kinase inhibitor 5, 6-dichloro-1-beta-D-ribofuranosylbenzimidazole induces nongenotoxic, DNA replication-independent

apoptosis of normal and leukemic cells, regardless of their p53 status. *BMC Cancer* **9**(281).

194. Minvielle-Sebastia, L., Preker, P.J., Wiederkehr, T., Strahm, Y., and Keller, W. 1997. The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. *Proc Natl Acad Sci U S A* **94**(15): 7897-7902.

195. Afonina, E., Stauber, R., and Pavlakis, G.N. 1998. The human poly(A)-binding protein 1 shuttles between the nucleus and the cytoplasm. *J Biol Chem* **273**(21): 13015-13021.

196. Lee, S., Neumann, M., Stearman, R., Stauber, R., Pause, A., Pavlakis, G.N., and Klausner, R.D. 1999. Transcription-dependent nuclear-cytoplasmic trafficking is required for the function of the von Hippel-Lindau tumor suppressor protein. *Mol Cell Biol* **19**(2): 1486-1497.

197. Lewis, M.D., and Roberts, B.J. 2003. Role of nuclear and cytoplasmic localization in the tumour-suppressor activity of the von Hippel-Lindau protein. *Oncogene* **22**(26): 3992-3997.

198. Strahl, B.D., and Allis, C.D. 2000. The language of covalent histone modifications. *Nature* **403**(6765): 41-45.

199. Jenuwein, T., and Allis, C.D. 2001. Translating the histone code. *Science* **293**(5532): 1074-1080.

200. Mariadason, J.M., Corner, G.A., and Augenlicht, L.H. 2000. Genetic reprogramming in pathways of colonic cell maturation induced by short chain fatty acids: comparison with trichostatin A, sulindac, and curcumin and implications for chemoprevention of colon cancer. *Cancer Res* **60**(16): 4561-4572.

201. Klampfer, L., Huang, J., Swaby, L.A., and Augenlicht, L. 2004. Requirement of histone deacetylase activity for signaling by STAT1. *J Biol Chem* **279**(29): 30358-30368.

202. Bode, K.A., Schroder, K., Hume, D.A., Ravasi, T, Heeg, K., Sweet, M.J., and Dalpke, A.H. 2007. Histone deacetylase inhibitors decrease Toll-like receptor-mediated activation of proinflammatory gene expression by impairing transcription factor recruitment. *Immunology* **122**(4): 596-606.

203. Juan, L.J., Shia, W.J., Chen, M.H., Yang, W.M., Seto, E., Lin, Y.S., and Wu, C.W. 2000. Histone deacetylases specifically down-regulate p53-dependent gene activation. *J Biol Chem* **275**(27): 20436-20443.

204. Martínez-Balbás, M.A., Bauer, U.M., Nielsen, S.J., Brehm, A., and Kouzarides, T. 2000. Regulation of E2F1 activity by acetylation. *EMBO J* **19**(4): 662-671.

205. Masumi, A., Yamakawa, Y., Fukazawa, H., Ozato, K., and Komuro, K. 2003. Interferon regulatory factor-2 regulates cell growth through its acetylation. *J Biol Chem* **278**(28): 25401-25407.

206. Thanos, D., and Maniatis, T. 1995. Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* **83**(7): 1091-1100.

207. Merika, M., Williams, A.J., Chen, G., Collins, T., and Thanos, D. 1998. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. . *Mol Cell* **1**: 277-287.

208. Chen, F.F., Jiang, G., Xu, K., and Zheng, J.N. 2013. Function and mechanism by which interferon regulatory factor-1 inhibits oncogenesis. *Oncol Lett* **5**(2): 417-423.

209. Parker, B., and Sukumar, S. 2003. Distant metastasis in breast cancer: molecular mechanisms and therapeutic targets. *Cancer Biol Ther* **2**(1): 14-21.

210. Moses, M.A. 1997. The Regulation of Neovascularization by Matrix Metalloproteinases and Their Inhibitors. *Stem cells* **15**: 180-189.

211. Tomanek, R.J., and Schatteman, G.C. 2000. Angiogenesis: new insights and therapeutic potential. *Anat Rec* **261**(3): 126-135.

212. Opdenakker, G., Van den Steen, P.E., Dubois, B., Nelissen, I., Van Coillie, E., Masure, S., Proost, P., and Van Damme, J. 2001. Gelatinase B functions as regulator and effector in leukocyte biology. *J Leukoc Biol* **69**(6): 851-859.

213. Tomanek, R.J., and Schatteman, G.C. 2000. Angiogenesis: new insights and therapeutic potential. *Anat Rec* **261**(3): 126-135.

214. McCawley, L.J., and Matrisian, L.M. 2000. Matrix metalloproteinases: multifunctional contributors to tumor progression. *Mol Med Today* **6**(4): 149-156.

215. Moses, M.A. 1997. The regulation of neovascularization of matrix metalloproteinases and their inhibitors. *Stem Cells* **15**(3): 180-189.

216. Janowska-Wieczorek, A., Marquez, L., Matsuzaki, A., Hashmi, H., Larratt, L., Boshkov, L., Turner, A., Zhang, M., Edwards, D. and Kossakowska, A. 1999. Expression of matrix metalloproteinases (MMP-2 and -9) and tissue inhibitors of metalloproteinases (TIMP-1 and -2) in acute myelogenous leukaemia blasts: comparison with normal bone marrow cells. *British Journal of Haematology* **105**(2): 402-411.

217. Kossakowska, A.E., Urbanski, S.J., and Janowska-Wieczorek, A. 2000. Matrix metalloproteinases and their tissue inhibitors - expression, role and regulation in human malignant non-Hodgkin's lymphomas. *Leuk Lymphoma* **39**(5-6): 485-493.

218. Roomi, M.W., Monterrey, J.C., Kalinovsky, T., Rath, M., and Niedzwiecki, A. 2009. Patterns of MMP-2 and MMP-9 expression in human cancer cell lines. *Oncol Rep* **21**(5): 1323-1333.

219. Huhtala, P., Chow, L.T., and Tryggvason, K. 1990. Structure of the human type IV collagenase gene. *J Biol Chem* **265**(19): 11077-11082.

220. Sato, H., Kita, M., and Seiki, M. 1993. v-Src activates the expression of 92-kDa type IV collagenase gene through the AP-1 site and the GT box homologous to retinoblastoma control elements. A mechanism regulating gene expression independent of that by inflammatory cytokines. *J Biol Chem* **268**(31): 23460-23468.

221. Sato, H., and Seiki, M. 1993. Regulatory mechanism of 92 kDa type IV collagenase gene expression which is associated with invasiveness of tumor cells. *Oncogene* **8**(2): 395-405.

222. Gum, R., Lengyel, E., Juarez, J., Chen, J.H., Sato, H., Seiki, M., and Boyd, D. 1996. Stimulation of 92-kDa gelatinase B promoter activity by ras is mitogen-activated protein kinase kinase 1-independent and requires multiple transcription factor binding sites including closely spaced PEA3/ets and AP-1 sequences. *J Biol Chem* **271**(18): 10672-10680.

223. Bond, M., Fabunmi, R.P., Baker, A.H., and Newby, A.C. 1998. Synergistic upregulation of metalloproteinase-9 by growth factors and inflammatory cytokines: an absolute requirement for transcription factor NF-kappa B. *FEBS Lett* **435**(1): 29-34.

224. Antoni, B.A., Stein, S.B., and Rabson, A.B. 1994. Regulation of human immunodeficiency virus infection: implications for pathogenesis. *Adv Virus Res* **43**: 53-145.

225. Roulston, A., Lin, R., Beauparlant, P., Wainberg, M.A., and Hiscott, J. 1995. Regulation of human immunodeficiency virus type 1 and cytokine gene expression in myeloid cells by NF-κB/Rel transcription factors. *Microbiol Rev* **59**: 481-505.

226. Cullen, B.R. 1991. Regulation of HIV-1 gene expression. *FASEB J* **5**: 2361-2368.

227. Tang, H., Kuhen, K.L., and F. Wong-Staal, F. 1999. Lentivirus replication and regulation. . *Annu Rev Genet* **33**: 133-170.

228. Van Lint, C., Ghysdael. J., Paras, P Jr., Burny, A., and Verdin, E. 1994. A transcriptional regulatory element is associated with a nuclease-hypersensitive site in the pol gene of human immunodeficiency virus type. . *J Virol* **68**: 2632-2648.

229. El Kharroubi, A., and Verdin, E. 1994. Protein-DNA interactions within DNase I-hypersensitive sites located downstream of the HIV-1 promoter. *J Biol Chem* **269**: 19916-19924.

230. Marsili, G., Borsetti, A., Sgarbanti, M., Remoli, A.L., Ridolfi, B., Stellacci, E., Ensoli, B., and Battistini, A. 2003. On the role of interferon regulatory factors in HIV-1 replication. *Ann N Y Acad Sci* **1010**: 29-42.

231. El Kharroubi, A., and Martin, M.A. . 1996. Cis-acting sequences located downstream of the human immunodeficiency virus type 1 promoter affect its chromatin structure and transcriptional activity. *Mol Cell Biol* **16**: 2958-2966.

232. Roebuck, K.A., Brenner, D.A., and Kagnoff, M.F. 1993. Identification of c-fos-responsive elements downstream of TAR in the long terminal repeat of human immunodeficiency virus type-1. . *J Clin Invest* **92**: 1336-1349.

233. Harada, H., Fujita, T., Miyamoto, M., Kimura, Y., Maruyama, M., Furia, A., Miyata, T., and Taniguchi, T. 1989. Structurally similar but functionally distinct factors, IRF-1 and IRF-2, bind to the same regulatory elements of IFN and IFN-inducible genes. *Cell* **58**: 729-739.

234. Verdin, E. 1991. DNase I-hypersensitive sites are associated with both long terminal repeats and with the intrangenic enhancer of integrated human immunodeficiency virus type 1. *J Virol* **65**: 6790-6799.

235. Sgarbanti, M., Borsetti, A., Moscufo, N., Bellocch,i M.C., Ridolfi, B., Nappi, F., Marsili, G., Marziali, G., Coccia, E.M., Ensoli, B., and Battistini, A. 2002. Modulation of human immunodeficiency virus 1 replication by interferon regulatory factors. *J Exp Med* **195**(10): 1359-1370.

236. Wu, Y., and Marsh, J.W. 2001. Selective transcription and modulation of resting T cell activity by preintegrated HIV DNA. . *Science* **293**: 1503-1506.

237. Jeang, K.T., and Gatignol, A. 1994. Comparison of regulatory features among primate lentiviruses. *Curr Top Microbiol Immunol* **188**: 123-144.

238. Chang, H.-K., Gallo, R.C., and Ensoli, B. 1995. Regulation of cellular gene expression and function by the human immunodeficiency virus type 1 Tat protein. *J Biomed Sci* **2**: 189-202.

239. Jones, K.A., and Peterlin, B.M. 1994. Control of RNA initiation and elongation at the HIV-1 promoter. . *Annu Rev Biochem* **63**: 717-743.

240. Cullen, B.R. 1992. Mechanism of action of regulatory proteins encoded by complex retroviruses. . *Microbiol Rev* **56**: 375-394.

241. Bohan, C.A., Kashanchi, F., Ensoli, B., Buonaguro, L., Boris-Lawrie, K.A., and Brady, J.N. . 1992. Analysis of Tat transactivation of human immunodeficiency virus transcription in vitro. *Gene Expr* **2**: 391-407.

242. Wang, I.M., Blanco, J.C., Tsai, M.J., and Ozato, K. 1996. Interferon regulatory factors and TFIIB cooperatively regulate interferon-responsive promoter activity in vivo and in vitro. *Mol Cell Biol* **16**: 6313-6324.

243. Veschambre, P., Roisin, A., and Jalinot, P. 1997. Biochemical and functional interaction of the human immunodeficiency virus type 1 Tat transactivator with the general transcription factor TFIIB. . *J Gen Virol* **78**: 2235-2245.

244. Hottiger, M.O., and Nabel, G.J. 1998. Interaction of human immunodeficiency virus type 1 Tat with the transcriptional coactivators p300 and CREB binding protein. *J Virol* **72**: 8252-8256.

245. Sgarbanti, M., Remoli, A.L., Marsili, G., Ridolfi, B., Borsetti, A., Perrotti, E., Orsatti, R., Ilari, R., Sernicola, L., Stellacci, E,, Ensoli, B., and Battistini, A. 2008. IRF-1 is required for full NF-kappaB transcriptional activity at the human

immunodeficiency virus type 1 long terminal repeat enhancer. *J Virol* **82**(7): 3632-3641.

246. Zhou, H., Xu, M., Huang, Q., Gates, A.T., Zhang, X.D., Castle, J.C., Stec, E., Ferrer, M., Strulovici, B., Hazuda, D.J., and Espeseth, A.S. 2008. Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* **4**(5): 495-504.

247. Allouch, A., Di Primio, C., Alpi, E., Lusic, M., Arosio, D., Giacca, M., and Cereseto, A. 2011. The TRIM family protein KAP1 inhibits HIV-1 integration. *Cell Host Microbe* **9**(6): 484-495.

248. Marsili, G., Remoli, A.L., Sgarbanti, M., and Battistini, A. 2004. Role of acetylases and deacetylase inhibitors in IRF-1-mediated HIV-1 long terminal repeat transcription. *Ann N Y Acad Sci* **1030**: 636-643.

249. Liang, C., Li, X., Quan, Y., Laughrea, M., Kleiman, L., Hiscott, J., and Wainberg, M.A. 1997. Sequence elements downstream of the human immunodeficiency virus type 1 long terminal repeat are required for efficient viral gene transcription. *J Mol Biol* **272**: 167-177.

250. Gorman, C.M., Moffat, L.F., and Howard, B.H. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. . *Mol Cell Biol* **2**: 1044-1051.

251. Battistini, A., Marsili, G., Sgarbanti, M., Ensoli, B., and Hiscott, J. 2002. IRF regulation of HIV-1 long terminal repeat activity. *J Interferon Cytokine Res* **22**(1): 27-37.

252. Aragane, Y., Schwarz, A., Luger, T.A., Ariizumi, K., Takashima, A. and Schwarz, T. . 1997. Ultraviolet light suppresses IFN-gamma-induced IL-7 gene expression in murine keratinocytes by interfering with IFN regulatory factors. *J Immunol* **158**: 5393-5399.

253. Li, B., Carey, M., and Workman, J.L. 2007. The role of chromatin during transcription. *Cell* **128**(4): 707-719.

254. Hebbes, T.R., Clayton, A.L., Thorne, A.W., and Crane-Robinson, C. . 1994. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. . *EMBO J* **13**: 1823-1830.

255. Braunstein, M., Rose, A.B., Holmes, S.G., Allis, C.D., and Broach, J.R. 1993. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. . *Genes Dev* **7**: 592-604.

256. Kouzarides, T. 2007. Chromatin modifications and their function. *Cell* **128**: 693-705.

257. Smith, E., and Shilatifard, A. . 2010. The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. *Mol Cell* **40**: 689-701.

258. Latchman, D.S. 1997. Transcription factors: an overview. *Int J Biochem Cell Biol* **29**: 1305-1312.

259. Shi, Y., Sawada, J., Sui, G., Affar, el B., Whetstine, J.R., Lan, F., Ogawa, H., Luke, M.P., and Nakatani, Y. . 2003. Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422**: 735-738.

260. Luo, R.X., Postigo, A.A., and Dean, D.C. 1998. Rb interacts with histone deacetylase to repress transcription. *Cell* **92**(4): 463-473.

261. Ross, J.F., Liu, X., and Dynlacht, B.D. 1999. Mechanism of transcriptional repression of E2F by the retinoblastoma tumor suppressor protein. *Mol Cell* **3**(2): 195-205.

262. Vandel, L., Nicolas, E., Vaute, O., Ferreira, R., Ait-Si-Ali, S., and Trouche, D. 2001. Transcriptional repression by the retinoblastoma protein through the recruitment of a histone methyltransferase. *Mol Cell Biol* **21**(19): 6484-6494.

263. Miyamoto, M., Fujita, T., Kimura, Y., Maruyama, M., Harada, H., et al. . 1988. Regulated expression of a gene encoding a nuclear factor, IRF-1, that specifically binds to IFN-beta gene regulatory elements. *Cell* **54**: 903-913.

264. Muhlethaler-Mottet, A., Di Berardino, W., Otten, L.A., and Mach, B. . 1998. Activation of the MHC class II transactivator CIITA by interferon-gamma requires cooperative interaction between Stat1 and USF-1. *Immunity* **8**: 157-166.

265. Ogasawara, K., Hida, S., Azimi, N., Tagaya, Y., Sato, T., et al. . 1998. Requirement for IRF-1 in the microenvironment supporting development of natural killer cells. *Nature* **391**: 700-703.

266. Salkowski, C.A., Kopydlowski, K., Blanco, J., Cody, M.J., McNally, R., et al. 1999. IL-12 is dysregulated in macrophages from IRF-1 and IRF-2 knockout mice. . *J Immunol* **163**: 1529-1536.

267. White, L.C., Wright, K.L., Felix, N.J., Ruffner, H., Reis, L.F., et al. 1996. Regulation of LMP2 and TAP1 genes by IRF-1 explains the paucity of CD8+ T cells in IRF-1-/- mice. *Immunity* **5**: 365-376.

268. Ruiz-Ruiz, C., Ruiz de Almodóvar, C., Rodríguez, A., Ortiz-Ferrón, G., Redondo, J.M., et al. 2004. The up-regulation of human caspase-8 by interferon-gamma in breast tumor cells requires the induction and action of the transcription factor interferon regulatory factor-1. *J Biol Chem* **279**: 19712-19720.

269. Gao, J., Senthil, M., Ren, B., Yan, J., Xing, Q., et al. . 2010. IRF-1 transcriptionally upregulates PUMA, which mediates the mitochondrial apoptotic pathway in IRF-1-induced apoptosis in cancer cells. . *Cell Death Differ* **17**: 699-709.

270. Wang, J., Zhang, W., Zhang, Y., Chen, Y., Zou, B., et al. 2009. c-Jun N-terminal kinase (JNK1) upregulates XIAP-associated factor 1 (XAF1) through interferon regulatory factor 1 (IRF-1) in gastrointestinal cancer. *Carcinogenesis* **30**: 222-229.

271. Clarke, N., Jimenez-Lara, A.M., Voltz, E., and Gronemeyer, H. . 2004. Tumor suppressor IRF-1 mediates retinoid and interferon anticancer signaling to death ligand TRAIL. *EMBO J* **23**: 3051-3060.

272. Frontini, M., Vijayakumar, M., Garvin, A., and Clarke, N. 2009. A ChIP-chip approach reveals a novel role for transcription factor IRF1 in the DNA damage response. *Nucleic Acids Res* **37**: 1073-1085.

273. Shi, L., Perin, J.C., Leipzig, J., Zhang, Z., and Sullivan, K.E. 2011. Genome-wide analysis of interferon regulatory factor I binding in primary human monocytes. *Gene* **487**(1): 21-28.

274. Rettino, A., and Clarke, N.M. 2013. Genome-wide Identification of IRF1 Binding Sites Reveals Extensive Occupancy at Cell Death Associated Genes. *J Carcinog Mutagen* **S6:009**: doi: 10.4172/2157-2518.S4176-4009.

275. Gilbert, N., and Allan, J. 2001. Distinctive higher-order chromatin structure at mammalian centromeres. *Proc Natl Acad Sci U S A* **98**(21): 11949-11954.

276. Gilbert, N., Boyle, S., Sutherland, H., de Las Heras, J., Allan, J., Jenuwein, T., and Bickmore, W.A. 2003. Formation of facultative heterochromatin in the absence of HP1. *EMBO J* **22**(20): 5540-5550.

277. Farnham, P.J. 2009. Insights from genomic profiling of transcription factors. *Nature Rev Genet* **10**(9): 605-616.

278. Solomon, M.J., Larsen, P. L., and Varshavsky, A. 1988. Mapping protein-DNA interactions *in vivo* with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* **53**: 937-947.

279. Blat, Y., and Kleckner, N. . 1999. Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell* **98**: 249-259.

280. Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T.L., Wilson, C.J., Bell, S.P., and Young, R.A. 2000. Genome-wide location and function of DNA binding proteins. *Science* **290**: 2306-2309.

281. Bentley, D.R. 2006. Whole-genome re-sequencing. *Curr Opin Genet Dev* **16**: 545-552.

282. Shendure, J., and Ji, H. . 2008. Next-generation DNA sequencing. *Nature Biotech* **26**: 1135-1145.

283. Mardis, E.R. 2008. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* **9**: 387-402.

284. Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. . *Science* **316**: 1497-1502.

285. Park, P.J. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**(10): 669-680.

286. Landt, S.G., *et al.* 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**(9): 1813-1831.

287. Edinburgh, U.o. 2009. The GenePool. In *Sequencing and Bioinformatics and the University of Edinburgh*.

288. ECO. 2007. Tech summary: Illumina's Solexa Sequencing Technology. In *SEQanswers*. Jelsoft Enterprises Ltd.

289. Illumina.com. 2014. Sequencing Technology, Sequencing by Synthesis, Illumina. In.

290. Brown, S.M. 2012. Sequencing-by-Synthesis: Explaining the Illumina Sequencing Technology. In *BitesizeBio*.

291. Illumina. 2010. Whole-Genome Chromatin IP Sequencing (ChIP-Seq). In *Data Sheet: Sequencing*. Illumina Inc, San Diego.

292. -. 2014. Sequencing Technology, Sequencing by Synthesis, Illumina. In. Illumina Inc., San Diego.

293. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996-1006.

294. Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Helmberg, W., et al. . 2005. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*: D39-D45.

295. Birney, E., Andrews, D., Bevan, P., Caccamo, M., Cameron, G., Chen, Y., Clarke, L., Coates, G., Cox, T., Cuff, J. et al. . 2004. Ensembl 2004. *Nucleic Acids Res* D468-D470.

296. Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(4): 1754-1760.

297. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**(9): R137.

298. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. . 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**(4): 576-589.

299. Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**: 1351-1359.

300. EpiGenie. 2013. EpiGenie Guide: ChIP Antibody Validation. In.

301. Bardet, A.F., He, Q., Zeitlinger, J., and Stark, A. 2011. A computational pipeline for comparative ChIP-seq analyses. *Nat Protoc* **7**(1): 45-61.

302. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**(5): 495-501.

303. Thomas-Chollier, M., Herrmann, C., Defrance, M., Sand, O., Thieffry, D., and van Helden, J. 2011. RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Research* **doi:10.1093**.

304. Thomas-Chollier, M., Darbo, E., Herrmann, C., Defrance, M., Thieffry, D., van Helden, J. 2012. A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs. . *Nat Protoc* **7**(8): 1551-1568.

305. Linzer, D.I., and Levine, A.J. 1979. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell* **17**: 43-52.

306. Lane, D.P., and Crawford, L.V. 1979. T antigen is bound to a host protein in SV40-transformed cells. *Nature* **278**: 261-263.

307. DeLeo, A.B., Jay, G., Appella, E., Dubois, G.C., Law, L.W., and Old, L.J. 1979. Detection of a transformation-related antigen in chemically induced sarcomas and other transformed cells of the mouse. *Proc Natl Acad Sci U S A* **76**: 2420-2424.

308. Vousden, K.H., and Lu, X. 2002. Live or let die: the cell's response to p53. . *Nat Rev Cancer* **2**: 594-604.

309. Levine, A.J., Hu, W., and Feng, Z. 2006. The P53 pathway: what questions remain to be explored? . *Cell Death Differ* **13**: 1027-1036.

310. Joerger, A.C., and Fersht, A.R. 2008. Structural biology of the tumor suppressor p53. *Annu Rev Biochem* **77**: 557-582.

311. Lozano, G., and Montes de Oca Luna, R. 1998. MDM2 function. *Biochim Biophys Acta* **1377**(2): M55-59.

312. Oren, M. 1999. Regulation of the p53 Tumor Suppressor Protein. *J Biol Chem* **274**: 36031-36034.

313. Landré, V., Pion, E., Narayan, V., Xirodimas, D.P., and Ball, K.L. 2013. DNA-binding regulates site-specific ubiquitination of IRF-1. *Biochem J* **449**(3): 707-717.

314. Zhan, Q., Chen, I.T., Antinore, M.J., and Fornace, A.J. Jr. 1998. Tumor suppressor p53 can participate in transcriptional induction of the GADD45 promoter in the absence of direct DNA binding. . *Cell Biol* **18**: 2768-2778.

315. Jin, S., Fan, F., Fan, W., Zhao, H., Tong, T., Blanck, P., Alomo, I., Rajasekaran, B., and Zhan, Q. 2001. Transcription factors Oct-1 and NF-YA regulate the p53-independent induction of the GADD45 following DNA damage. . *Oncogene* **20**: 2683-2690.

316. Tompa, M., Li, N., Bailey, T., Church, G., De Moor, B., Eskin, E., Favorov, A., Frith, M., Fu, Y., Kent, W., et al. . 2005. Assessing computational tools for the discovery of transcription factor binding sites. . *Nat Biotechnol* **23**(1): 137-144.

317. Won, K.J., Ren, B., and Wang, W. 2010. Genome-wide prediction of transcription factor binding sites using an integrated model. *Genome Biol* **11**(1): R7.

318. Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. 2011. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res* **21**(3): 447-455.

319. Wingender, E., Dietze, P., Karas, H., and Knüppel, R. . 1996. TRANSFAC: A database on transcription factors and their DNA binding sites. *Nucleic Acids Res* **24**: 238-241.

320. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W., and Lenhard, B. . 2004. JASPAR: An open-access database for eukaryotic transcription factor binding profiles. . *Nucleic Acids Res* **32**: D91-D94.

321. Boyle, A., Davis, S., Shulha, H., Meltzer, P., Margulies, E., Weng, Z., Furey, T., and Crawford, G. . 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**: 311-322.

322. Gaulton, K., Nammo, T., Pasquali, L., Simon, J., Giresi, P., Fogarty, M., Panhuis, T., Mieczkowski, P., Secchi, A., Bosco, D., et al. . 2010. A map of open chromatin in human pancreatic islets. *Nat Genet* **42**: 255-259.

323. Fu, Y., Sinha, M., Peterson, C., and Weng, Z. . 2008. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet* **4**: e1000138.

324. Hesselberth, J., Chen, X., Zhang, Z., Sabo, P., Sandstrom, R., Reynolds, A., Thurman, R., Neph, S., Kuehn, M., Noble, W., et al. 2009. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. . *Nat Methods* **6**: 283-289.

325. Chen, X., Hoffman, M.M., Bilmes, J.A., Hesselberth, J.R., and Noble, W.S. 2010. A dynamic Bayesian network for identifying protein-binding footprints from single molecule-based sequencing data. *Bioinformatics* **26**: i334-i342.

326. Ernst, J., Plasterer, H., Simon, I., and Bar-Joseph, Z. . 2010. Integrating multiple evidence sources to predict transcription factor binding in the human genome. . *Genome Res* **20**: 526-536.

327. Visel, A., Blow, M., Li, Z., Zhang, T., Akiyama, J., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. . 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854-858.

328. Galas, D., and Schmitz, A. 1978. DNAse footprinting a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res* **5**: 3157-3170.

329. Bulger, M., and Groudine, M. 2011. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**(3): 327-339.

330. Galon, J., Sudarshan, C., Ito, S., Finbloom, D., and O'Shea, J.J. 1999. IL-12 induces IFN regulating factor-1 (IRF-1) gene expression in human NK and T cells. *J Immunol* **162**(12): 7256-7262.

331. Human Genome Sequencing, C. 2004. Finishing the euchromatic sequence of the human genome. *Nature* **431**(7011): 931-945.

332. Hertveldt, K., Belien, T., and Volckaert, G. . 2009. General M13 phage display: M13 phage display in identification and characterization of protein-protein interactions. *Methods Mol Biol* **502**: 321-339.

333. Molek, P., Strukelj, B., and Bratkovic, T. 2011. Peptide Phage Display as a Tool for Drug Discovery: Targeting Membrane Receptors *Molecules* **16**: 857-887.

334. De Berardinis, P., and Haigwood, N. . 2004. New recombinant vaccines based on the use of prokaryotic antigen-display systems. *Expert Rev* **3**: 673.

335. Sergeeva, A., Kolonin, M.G., Molldrem, J.J., Pasqualini, R., and Arap, W. . 2006. Display technologies: application for the discovery of drug and gene delivery agents. *Adv Drug Deliv Rev* **58**: 1622-1654.

336. Smith, G.P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**(4705): 1315-1317.

337. Smith, G.P., and Petrenko, V.A. . 1997. Phage Display. *Chem Rev* **97**: 391-410.

338. Bratkovič, T. 2010. Progress in phage display: evolution of the technique and its applications. *Mol Life Sci* **67**: 749-767.

339. Phagedisplay.net. 2010. Phage Display Innovative solutions. In.

340. Koivunen, E., Arap, W., Rajotte, D., Lahdenranta, J., and Pasqualini, R. 1999. Identification of receptor ligands with phage display peptide libraries. *J Nucl Med* **40**(5): 883-888.

341. Fukunaga, K., and Taki, M. 2012. Practical Tips for Construction of Custom Peptide Libraries and Affinity Selection by Using Commercially Available Phage Display Cloning Systems. *J Nucleic Acids* **2012**(295719): 9.

342. Matochko, W.L., Chu, K., Jin, B., Lee, S.W., Whitesides, G.M., and Derda, R. 2012. Deep sequencing analysis of phage libraries using Illumina platform. *Methods* **58**(1): 47-55.

343. Dias-Neto, E., Nunes, D.N., Giordano, R.J., Sun, J., Botz, G.H., Yang, K., Setubal, J.C., Pasqualini, R., and Arap, W. 2009. Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis. *PLoS One* **4**(12): e8338.

344. Ernst, A., Gfeller, D., Kan, Z., Seshagiri, S., Kim, P.M., Bader, G.D., and Sidhu, S.S. 2010. Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol Biosyst* **6**(10): 1782-1790.

345. Ravn, U., Gueneau, F., Baerlocher, L., Osteras, M., Desmurs, M., Malinge, P., Magistrelli, G., Farinelli, L., Kosco-Vilbois, M.H., and Fischer, N. 2010. By-passing in vitro screening--next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic Acids Res* **38**(21): e193.

346. 't Hoen, P.A., Jirka, S.M., Ten Broeke, B.R., Schultes, E.A., Aguilera, B., Pang, K.H., Heemskerk, H., Aartsma-Rus, A., van Ommen, G.J., and den Dunnen, J.T. 2012. Phage display screening without repetitious selection rounds. *Anal Biochem* **421**(2): 622-631.

347. Ru, B., Huang, J., Dai, P., Li, S., Xia, Z., Ding, H., Lin, H., Guo, F., and Wang, X. 2010. MimoDB: a new repository for mimotope data derived from phage display technology. *Molecules* **15**(11): 8279-8288.

348. Derda, R., Tang, S.K., Li, S.C., Ng, S., Matochko, W., and Jafari, M.R. 2011. Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* **16**(2): 1776-1803.

349. NEB. 2014. Panning with a Ph.D. Phage Display Library. In (ed. https://www.neb.com/products/e8120-phd-c7c-phage-display-peptide-library-kit).

350. Illumina. 2011. Quality scores for next-generation sequencing: Assessing sequencing accuracy using Phred quality scoring. In. Illumina Inc, San Diego.

351. Matochko, W.L., Cory Li, S., Tang, S.K., and Derda, R. 2013. Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Research* **42**(3): 1784-1798.

352. Bailey, T.L., and Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**: 28-36.

353. Scanprosite. 2014. SIB Swiss Institute for Bioinformatics. In *http://prositeexpasyorg/scanprosite/*.

354. Thomas, P.D., Kejariwal, A., Guo, N., Mi, H., Campbell, M.J., Muruganujan, A., and Lazareva-Ulitsky, B. 2006. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res* **34**(Web server issue): W645-650.

355. Komiya, Y., and Habas, R. . 2008. Wnt signal transduction pathways. *Organogenesis* **4**(2): 68-75.

356. Pace, N.J., and Weerapana, E. 2014. Zinc-Binding Cysteines: Diverse Functions and Structural Motifs. *Biomolecules* **4**: 419-434.

357. Nielsen, A.L., Jørgensen, P., Lerouge, T., Cerviño, M., Chambon, P., and Losson, R. 2004. Nizp1, a novel multitype zinc finger protein that interacts with the

NSD1 histone lysine methyltransferase through a unique C2HR motif. *Mol Cell Biol* **24**(12): 5184-5196.

358. Borden, K.L., and Freemont, P.S. 1996. The RING finger domain: a recent example of a sequence-structure family. *Curr Opin Struct Biol* **6**(3): 395-401.

359. Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. 2013. The architecture of human general transcription factor TFIID core complex. *Nature* **493**(7434)): 699-702.

Appendix I: Recombinant ZNF350 binds to the Mf1 and Mf2 domains of IRF-1

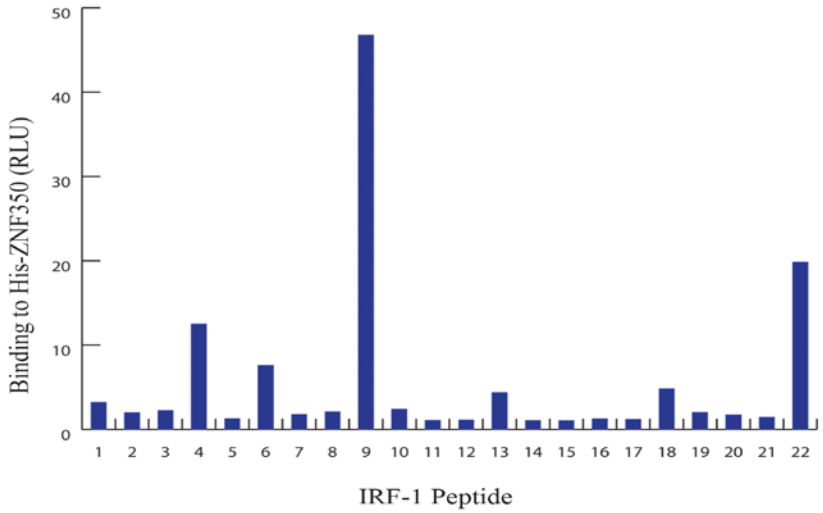


Figure 6-1: His-ZNF350 binds to the Mf1 and Mf2 domains of IRF-1

IRF-1 peptides were immobilised on a microtitre plate then His-ZNF350 was added and binding detected using an anti-His antibody and enhanced chemiluminescence.

Appendix II: ZNF350 biological replicate 1 peaks generated by ChIP-seq

Chromo	*Start*	*End*	*Peak Identifier*	*Peak Size*
chr1	446204	446341	MACS_peak_1	58.93
chr1	13471299	13471447	MACS_peak_2	56
chr1	26406104	26406276	MACS_peak_3	65.37
chr1	46178591	46178722	MACS_peak_4	59.88
chr1	47440451	47440592	MACS_peak_5	57.54
chr1	52700368	52700489	MACS_peak_6	56.26
chr1	62968051	62968223	MACS_peak_7	54.69
chr1	64953560	64953661	MACS_peak_8	68.77
chr1	68661742	68661878	MACS_peak_9	51.24
chr1	70218600	70218813	MACS_peak_10	55.1
chr1	70396324	70396472	MACS_peak_11	50.3
chr1	83778009	83778157	MACS_peak_12	53.97
chr1	98517146	98517282	MACS_peak_13	52.93
chr1	100526849	100526993	MACS_peak_14	52.56
chr1	118694840	118694987	MACS_peak_15	54.38
chr1	119398708	119398855	MACS_peak_16	54.58
chr1	121351914	121352466	MACS_peak_17	111.71
chr1	121352491	121352727	MACS_peak_18	68.74
chr1	121354118	121354610	MACS_peak_19	85.37
chr1	121354705	121355290	MACS_peak_20	139.27
chr1	121355696	121355845	MACS_peak_21	175.7
chr1	121355900	121356195	MACS_peak_22	64.21
chr1	121356261	121356477	MACS_peak_23	50.79
chr1	121356887	121357148	MACS_peak_24	50.75
chr1	121357742	121358123	MACS_peak_25	98.62
chr1	121359006	121359220	MACS_peak_26	82.41
chr1	121360004	121360273	MACS_peak_27	80.39
chr1	121365721	121366009	MACS_peak_28	70.91
chr1	121386245	121386692	MACS_peak_29	55.99
chr1	121414439	121414506	MACS_peak_30	60.16
chr1	121458715	121458918	MACS_peak_31	86.93
chr1	121462524	121462620	MACS_peak_32	58.83
chr1	121468055	121468240	MACS_peak_33	82.62
chr1	121478605	121478765	MACS_peak_34	58.73
chr1	121479249	121479872	MACS_peak_35	106.45
chr1	121480087	121480619	MACS_peak_36	82.33
chr1	121482162	121482536	MACS_peak_37	58.61
chr1	121482997	121483668	MACS_peak_38	61.54

chr1	121484038	121485438	MACS_peak_39	2751.35
chr1	142614105	142614227	MACS_peak_40	51.03
chr1	142946939	142947115	MACS_peak_41	56.1
chr1	144583961	144584083	MACS_peak_42	55.63
chr1	145444306	145444431	MACS_peak_43	51.05
chr1	148256772	148256916	MACS_peak_44	54.03
chr1	148324528	148324747	MACS_peak_45	69.79
chr1	148823374	148823503	MACS_peak_46	54.33
chr1	154563061	154563206	MACS_peak_47	50.93
chr1	156283800	156283943	MACS_peak_48	55.42
chr1	164182182	164182318	MACS_peak_49	52.58
chr1	168133931	168134080	MACS_peak_50	51.5
chr1	170136596	170136744	MACS_peak_51	55.39
chr1	171068118	171068325	MACS_peak_52	69.78
chr1	174323746	174323960	MACS_peak_53	54.52
chr1	194555184	194555327	MACS_peak_54	52.42
chr1	195869300	195869414	MACS_peak_55	52.6
chr1	198527722	198527866	MACS_peak_56	53.27
chr1	201728203	201728330	MACS_peak_57	53.21
chr1	206710675	206710813	MACS_peak_58	58.22
chr1	213736885	213736979	MACS_peak_59	53.55
chr1	213991439	213991567	MACS_peak_60	63.17
chr1	215687349	215687456	MACS_peak_61	54.01
chr1	221419150	221419266	MACS_peak_62	51.07
chr1	221602959	221603109	MACS_peak_63	53.12
chr1	227704716	227704809	MACS_peak_64	58.84
chr1	228485308	228485441	MACS_peak_65	69.34
chr1	228597651	228597754	MACS_peak_66	68.13
chr1	229192730	229192859	MACS_peak_67	57.72
chr1	232105865	232106024	MACS_peak_68	53.42
chr1	236876911	236878363	MACS_peak_69	91.45
chr1	238965308	238965467	MACS_peak_70	56.42
chr1	242328203	242328376	MACS_peak_71	50.87
chr1	247207886	247208011	MACS_peak_72	55.78
chr1	248457512	248457659	MACS_peak_73	60.19
chr10	753732	753878	MACS_peak_74	54.02
chr10	4904069	4904208	MACS_peak_75	58.45
chr10	7729670	7729780	MACS_peak_76	54.72
chr10	21715294	21715386	MACS_peak_77	69.09
chr10	38778067	38778174	MACS_peak_78	75.77
chr10	38886901	38887188	MACS_peak_79	50.42
chr10	39047619	39047753	MACS_peak_80	50.27
chr10	39081542	39081664	MACS_peak_81	54.79

chr10	39101958	39102046	MACS_peak_82	79.18
chr10	39105187	39105369	MACS_peak_83	51.63
chr10	39111161	39111336	MACS_peak_84	54.79
chr10	39121378	39121482	MACS_peak_85	63
chr10	39128752	39128947	MACS_peak_86	64.47
chr10	39135540	39135605	MACS_peak_87	59.49
chr10	39138066	39138223	MACS_peak_88	93.71
chr10	42356927	42357245	MACS_peak_89	120.44
chr10	42358550	42360099	MACS_peak_90	52.98
chr10	42360529	42361320	MACS_peak_91	110.06
chr10	42364295	42364955	MACS_peak_92	128.39
chr10	42364957	42365274	MACS_peak_93	79.4
chr10	42365291	42366104	MACS_peak_94	170.87
chr10	42366147	42366549	MACS_peak_95	97.87
chr10	42369121	42369871	MACS_peak_96	84.33
chr10	42370074	42370970	MACS_peak_97	70.52
chr10	42371043	42371362	MACS_peak_98	55.23
chr10	42374115	42375020	MACS_peak_99	57.36
chr10	42378507	42378655	MACS_peak_100	56.43
chr10	42380028	42380776	MACS_peak_101	152.97
chr10	42383235	42383903	MACS_peak_102	397.42
chr10	42384166	42384863	MACS_peak_103	362.46
chr10	42384896	42385732	MACS_peak_104	1476.78
chr10	42385788	42386646	MACS_peak_105	378.94
chr10	42386921	42389624	MACS_peak_106	2417.98
chr10	42390248	42392000	MACS_peak_107	1010.49
chr10	42392011	42394664	MACS_peak_108	1641.53
chr10	42394670	42395304	MACS_peak_109	162.31
chr10	42395362	42397360	MACS_peak_110	1320.63
chr10	42398303	42399747	MACS_peak_111	718.23
chr10	42400092	42400801	MACS_peak_112	895.15
chr10	42402220	42402415	MACS_peak_113	53.88
chr10	42404269	42404744	MACS_peak_114	62.49
chr10	42406710	42406810	MACS_peak_115	118.51
chr10	42408215	42408404	MACS_peak_116	123.41
chr10	42409183	42409431	MACS_peak_117	149.4
chr10	42527381	42527622	MACS_peak_118	74.65
chr10	42527894	42528156	MACS_peak_119	242.16
chr10	42529261	42530399	MACS_peak_120	654.4
chr10	42531456	42531552	MACS_peak_121	51.27
chr10	42531845	42531975	MACS_peak_122	56.13
chr10	42532362	42532627	MACS_peak_123	96.36
chr10	42533780	42533935	MACS_peak_124	185.71

chr10	42534453	42534784	MACS_peak_125	183.37
chr10	42535084	42535288	MACS_peak_126	122.64
chr10	42535331	42535572	MACS_peak_127	66.25
chr10	42535682	42536099	MACS_peak_128	88.19
chr10	42536551	42537004	MACS_peak_129	230.01
chr10	42537404	42537870	MACS_peak_130	89.47
chr10	42538537	42538712	MACS_peak_131	135.61
chr10	42538797	42539572	MACS_peak_132	55.53
chr10	42539582	42540352	MACS_peak_133	207.99
chr10	42540435	42540904	MACS_peak_134	135.36
chr10	42541298	42541991	MACS_peak_135	198.36
chr10	42542722	42543179	MACS_peak_136	176.58
chr10	42543543	42543934	MACS_peak_137	110.45
chr10	42544356	42544836	MACS_peak_138	199.26
chr10	42545177	42545280	MACS_peak_139	79.05
chr10	42545994	42546163	MACS_peak_140	180.1
chr10	42596692	42597216	MACS_peak_141	820.36
chr10	42597307	42599002	MACS_peak_142	746.38
chr10	42599014	42599441	MACS_peak_143	405.2
chr10	42599491	42600627	MACS_peak_144	1155.21
chr10	86191062	86191190	MACS_peak_145	70.02
chr10	93095852	93095966	MACS_peak_146	56.78
chr10	93718912	93719040	MACS_peak_147	71.68
chr10	100590784	100590928	MACS_peak_148	51.49
chr10	109239891	109240029	MACS_peak_149	58.22
chr10	112004539	112004655	MACS_peak_150	50.25
chr10	113319593	113319741	MACS_peak_151	64.12
chr10	124656844	124656990	MACS_peak_152	50.72
chr10	135500616	135501267	MACS_peak_153	67.41
chr11	9496148	9496284	MACS_peak_154	59.16
chr11	24160532	24160679	MACS_peak_155	56.22
chr11	29422031	29422149	MACS_peak_156	63.72
chr11	35270264	35270358	MACS_peak_157	57.35
chr11	36916242	36916387	MACS_peak_158	54.42
chr11	37769479	37769631	MACS_peak_159	61.44
chr11	48740488	48740599	MACS_peak_160	54.56
chr11	48810198	48810360	MACS_peak_161	94.06
chr11	49048797	49048938	MACS_peak_162	58.01
chr11	50637366	50637530	MACS_peak_163	50.81
chr11	50726002	50726177	MACS_peak_164	51.31
chr11	51567203	51567404	MACS_peak_165	54.44
chr11	51568507	51568627	MACS_peak_166	50.73
chr11	51568915	51569005	MACS_peak_167	88.59

chr11	51569588	51569843	MACS_peak_168	105.45
chr11	51569908	51570360	MACS_peak_169	108.26
chr11	51572048	51572221	MACS_peak_170	72.27
chr11	51572439	51572718	MACS_peak_171	53.07
chr11	51573086	51573199	MACS_peak_172	68.54
chr11	51576921	51577033	MACS_peak_173	66.93
chr11	51578567	51578831	MACS_peak_174	85.37
chr11	51579445	51579701	MACS_peak_175	153.54
chr11	51579956	51580372	MACS_peak_176	59.94
chr11	51580391	51580741	MACS_peak_177	59.65
chr11	51580764	51581127	MACS_peak_178	206.17
chr11	51582416	51583713	MACS_peak_179	297.92
chr11	51584161	51585134	MACS_peak_180	551.4
chr11	51586228	51586643	MACS_peak_181	129.8
chr11	51586655	51587458	MACS_peak_182	668.36
chr11	51588234	51588706	MACS_peak_183	121.95
chr11	51588720	51589536	MACS_peak_184	546.24
chr11	51590286	51590829	MACS_peak_185	157.08
chr11	51590836	51591249	MACS_peak_186	444.68
chr11	51592429	51593351	MACS_peak_187	176.68
chr11	68013529	68013668	MACS_peak_188	55.94
chr11	74441329	74441473	MACS_peak_189	57.35
chr11	76593735	76593856	MACS_peak_190	50.32
chr11	77229223	77229354	MACS_peak_191	65.08
chr11	77537754	77537887	MACS_peak_192	52.94
chr11	94369819	94369966	MACS_peak_193	51.66
chr11	94978016	94978172	MACS_peak_194	55.48
chr11	96753419	96753549	MACS_peak_195	58.47
chr11	100982928	100983056	MACS_peak_196	57.39
chr11	101550075	101550215	MACS_peak_197	50.34
chr11	112909130	112909261	MACS_peak_198	55.9
chr11	121759450	121759560	MACS_peak_199	53.2
chr11	127398442	127398577	MACS_peak_200	59.13
chr11	127651231	127651374	MACS_peak_201	56.88
chr11	131100059	131100216	MACS_peak_202	64.86
chr12	8504534	8504679	MACS_peak_203	53.46
chr12	16723494	16723640	MACS_peak_204	52.13
chr12	17709495	17709619	MACS_peak_205	55.83
chr12	18522749	18522872	MACS_peak_206	69.45
chr12	22047779	22047915	MACS_peak_207	54.17
chr12	24284985	24285111	MACS_peak_208	56.22
chr12	32558431	32558548	MACS_peak_209	61.84
chr12	33397352	33397487	MACS_peak_210	59.4

chr12	34163714	34163848	MACS_peak_211	57.3
chr12	34842087	34842180	MACS_peak_212	52.94
chr12	34846889	34847030	MACS_peak_213	66.37
chr12	34853800	34853955	MACS_peak_214	85.1
chr12	38002222	38002298	MACS_peak_215	57.86
chr12	38022523	38022587	MACS_peak_216	64.15
chr12	38036675	38036865	MACS_peak_217	59.17
chr12	38196323	38196464	MACS_peak_218	50.3
chr12	57197103	57197204	MACS_peak_219	66.18
chr12	65976586	65976783	MACS_peak_220	63.09
chr12	69119320	69119467	MACS_peak_221	53.99
chr12	69222558	69222709	MACS_peak_222	65.72
chr12	69229617	69229765	MACS_peak_223	56
chr12	69233294	69233440	MACS_peak_224	53.62
chr12	78466738	78466885	MACS_peak_225	54.18
chr12	81307338	81307482	MACS_peak_226	54.44
chr12	83697696	83697832	MACS_peak_227	55.78
chr12	90880058	90880239	MACS_peak_228	52.83
chr12	96258794	96258933	MACS_peak_229	52.95
chr12	99709747	99709919	MACS_peak_230	56.27
chr12	105641689	105641829	MACS_peak_231	54.75
chr12	109980254	109980386	MACS_peak_232	59.32
chr12	116472402	116472547	MACS_peak_233	54.61
chr12	116877092	116877223	MACS_peak_234	59.32
chr12	118271902	118272046	MACS_peak_235	57.35
chr13	21333333	21333467	MACS_peak_236	57.3
chr13	22740171	22740317	MACS_peak_237	56.86
chr13	28623348	28623497	MACS_peak_238	51.32
chr13	29085919	29086057	MACS_peak_239	59.95
chr13	29821152	29821287	MACS_peak_240	54.95
chr13	33323075	33323208	MACS_peak_241	55.42
chr13	35858483	35858617	MACS_peak_242	51.7
chr13	37749645	37749792	MACS_peak_243	56.69
chr13	40370510	40370571	MACS_peak_244	53.06
chr13	42006339	42006471	MACS_peak_245	60.12
chr13	58349120	58349267	MACS_peak_246	51.2
chr13	70043105	70043252	MACS_peak_247	53.6
chr13	76827899	76828046	MACS_peak_248	55.8
chr13	92078252	92078384	MACS_peak_249	59.57
chr13	101594105	101594254	MACS_peak_250	56.26
chr13	103534208	103534320	MACS_peak_251	59.06
chr13	107091994	107092138	MACS_peak_252	52.51
chr13	107830615	107830761	MACS_peak_253	54.7

chr13	108892750	108892855	MACS_peak_254	56.67
chr13	110325970	110326107	MACS_peak_255	55.78
chr13	111895878	111896020	MACS_peak_256	55.99
chr13	112147948	112148076	MACS_peak_257	68.89
chr13	113347261	113347406	MACS_peak_258	52.71
chr13	113917186	113917375	MACS_peak_259	53.88
chr14	19038339	19038534	MACS_peak_260	64.47
chr14	30545267	30545373	MACS_peak_261	67.19
chr14	34746465	34746623	MACS_peak_262	64.63
chr14	43897547	43897701	MACS_peak_263	55.78
chr14	70853956	70854101	MACS_peak_264	57.13
chr14	75662704	75662851	MACS_peak_265	56.69
chr14	77113558	77113707	MACS_peak_266	56.26
chr15	21313042	21313176	MACS_peak_267	59.32
chr15	23479916	23480011	MACS_peak_268	57.68
chr15	29491248	29491378	MACS_peak_269	54.34
chr15	32746037	32746184	MACS_peak_270	56.69
chr15	37806238	37806363	MACS_peak_271	55.58
chr15	49840271	49840420	MACS_peak_272	53.76
chr15	51867206	51867336	MACS_peak_273	56.14
chr15	77648952	77649096	MACS_peak_274	53.11
chr15	84372489	84372620	MACS_peak_275	60.36
chr15	84887661	84887794	MACS_peak_276	50.34
chr15	86757582	86757697	MACS_peak_277	73.69
chr16	16554924	16555057	MACS_peak_278	59.88
chr16	21889806	21889904	MACS_peak_279	56
chr16	22045991	22046095	MACS_peak_280	59.32
chr16	29483738	29483877	MACS_peak_281	50.24
chr16	32154489	32154693	MACS_peak_282	56.82
chr16	32791898	32792169	MACS_peak_283	64.82
chr16	33872897	33873069	MACS_peak_284	83.07
chr16	33878542	33878613	MACS_peak_285	50.6
chr16	33879645	33879706	MACS_peak_286	58.78
chr16	34002045	34002262	MACS_peak_287	50.7
chr16	35229037	35229318	MACS_peak_288	55.96
chr16	46385800	46402363	MACS_peak_289	3100
chr16	46402398	46402671	MACS_peak_290	110.91
chr16	46402681	46409531	MACS_peak_291	3100
chr16	46414176	46414545	MACS_peak_292	246.33
chr16	46418945	46422486	MACS_peak_293	1656.76
chr16	46422517	46426551	MACS_peak_294	1517.06
chr16	46426919	46431530	MACS_peak_295	142.75
chr16	46432048	46435582	MACS_peak_296	2317.51

chr16	46449571	46449831	MACS_peak_297	100.32
chr16	46455761	46455959	MACS_peak_298	50.09
chr16	58850285	58850350	MACS_peak_299	68.48
chr16	58892540	58892678	MACS_peak_300	50.79
chr16	64918121	64918263	MACS_peak_301	55.66
chr16	66483871	66484007	MACS_peak_302	64.98
chr16	69280034	69280182	MACS_peak_303	51.89
chr16	71050268	71050427	MACS_peak_304	52.09
chr16	71724071	71724210	MACS_peak_305	56.14
chr16	74421986	74422131	MACS_peak_306	52.71
chr16	84718250	84718395	MACS_peak_307	56.65
chr16	85087639	85087786	MACS_peak_308	56.22
chr16	90183158	90183376	MACS_peak_309	54.53
chr17	7572813	7572970	MACS_peak_310	85.99
chr17	15458553	15458700	MACS_peak_311	53.79
chr17	18521127	18521261	MACS_peak_312	55.37
chr17	22244690	22246934	MACS_peak_313	536.92
chr17	22246992	22248927	MACS_peak_314	247.56
chr17	22248950	22249857	MACS_peak_315	258.92
chr17	22249914	22252382	MACS_peak_316	813.19
chr17	22252394	22258891	MACS_peak_317	706.69
chr17	22258905	22261711	MACS_peak_318	1178.01
chr17	22261802	22262214	MACS_peak_319	56.52
chr17	22262638	22263006	MACS_peak_320	173.47
chr17	25280029	25280231	MACS_peak_321	58.01
chr17	26406945	26407081	MACS_peak_322	51.24
chr17	27528513	27528656	MACS_peak_323	68.21
chr17	29127575	29127664	MACS_peak_324	53.55
chr17	31252414	31252523	MACS_peak_325	65.37
chr17	31327280	31327424	MACS_peak_326	51.15
chr17	51408452	51408603	MACS_peak_327	58.09
chr17	51882244	51882373	MACS_peak_328	51.29
chr17	57943476	57943597	MACS_peak_329	62.91
chr17	62712898	62713038	MACS_peak_330	64.58
chr17	80254149	80254267	MACS_peak_331	75.19
chr18	76658	76801	MACS_peak_332	57.57
chr18	107842	109414	MACS_peak_333	411.99
chr18	109609	110062	MACS_peak_334	83.75
chr18	15174023	15174389	MACS_peak_335	69.24
chr18	18512817	18513416	MACS_peak_336	181.06
chr18	18515682	18516466	MACS_peak_337	217.34
chr18	18516518	18516690	MACS_peak_338	101.33
chr18	18517083	18518058	MACS_peak_339	199.39

chr18	18518078	18518241	MACS_peak_340	127.26
chr18	18518319	18518690	MACS_peak_341	64.65
chr18	18518700	18519229	MACS_peak_342	305.93
chr18	18519245	18520343	MACS_peak_343	461.98
chr18	18600704	18600850	MACS_peak_344	56.91
chr18	25133124	25133265	MACS_peak_345	58.01
chr18	54691872	54691993	MACS_peak_346	53.88
chr19	10584489	10584655	MACS_peak_347	68.91
chr19	14237349	14237560	MACS_peak_348	65.4
chr19	15795113	15795247	MACS_peak_349	62.69
chr19	24605636	24605972	MACS_peak_350	50.38
chr19	24619707	24620299	MACS_peak_351	61
chr19	24628728	24629277	MACS_peak_352	64.53
chr19	27731787	27732812	MACS_peak_353	1064.82
chr19	27732937	27733654	MACS_peak_354	583.09
chr19	27733781	27734899	MACS_peak_355	525.83
chr19	27736102	27736963	MACS_peak_356	734.74
chr19	27737354	27738752	MACS_peak_357	763.03
chr19	27738822	27739009	MACS_peak_358	101.64
chr19	27739094	27739532	MACS_peak_359	192.35
chr19	27740005	27740320	MACS_peak_360	117.09
chr19	27789615	27789746	MACS_peak_361	79.87
chr19	31253674	31253816	MACS_peak_362	61.02
chr19	44879011	44879133	MACS_peak_363	53.03
chr19	48005426	48005563	MACS_peak_364	69.76
chr19	50623115	50623329	MACS_peak_365	61.82
chr19	52468071	52469469	MACS_peak_366	1356.7
chr19	52471827	52471916	MACS_peak_367	134.16
chr19	52472269	52472430	MACS_peak_368	96.9
chr19_gl000208_random	38879	39184	MACS_peak_369	59.8
chr2	14167418	14167563	MACS_peak_370	51.28
chr2	22971461	22971607	MACS_peak_371	55.19
chr2	47644240	47644379	MACS_peak_372	55.75
chr2	48524848	48524976	MACS_peak_373	52.81
chr2	49036388	49036524	MACS_peak_374	50.28
chr2	59270406	59270553	MACS_peak_375	53.41
chr2	74026522	74026671	MACS_peak_376	52.41
chr2	78983315	78983448	MACS_peak_377	59.32
chr2	83295468	83295611	MACS_peak_378	53.52
chr2	85225486	85225624	MACS_peak_379	56.17
chr2	85782616	85782763	MACS_peak_380	56.64
chr2	89850307	89850861	MACS_peak_381	59.32
chr2	89853136	89853372	MACS_peak_382	93.71

chr2	89866755	89867593	MACS_peak_383	145.51
chr2	89867893	89868980	MACS_peak_384	146.35
chr2	89869733	89870122	MACS_peak_385	82.69
chr2	89870364	89870544	MACS_peak_386	73.72
chr2	89871078	89871323	MACS_peak_387	73.58
chr2	89872067	89872456	MACS_peak_388	67.88
chr2	89872605	89873075	MACS_peak_389	60.23
chr2	89875023	89875481	MACS_peak_390	187.29
chr2	89878329	89878657	MACS_peak_391	58.95
chr2	89878926	89879393	MACS_peak_392	106.36
chr2	89879741	89879994	MACS_peak_393	80.98
chr2	90375193	90376668	MACS_peak_394	54.49
chr2	90381141	90381544	MACS_peak_395	51.3
chr2	90387776	90388245	MACS_peak_396	123.48
chr2	90389217	90390638	MACS_peak_397	75.11
chr2	90392016	90392803	MACS_peak_398	85.83
chr2	90429858	90430122	MACS_peak_399	51.59
chr2	91608546	91609171	MACS_peak_400	51.75
chr2	91609254	91609733	MACS_peak_401	61.82
chr2	91610579	91612194	MACS_peak_402	75.92
chr2	91614317	91614946	MACS_peak_403	69.79
chr2	91768640	91769124	MACS_peak_404	50.6
chr2	91888996	91889236	MACS_peak_405	62.07
chr2	91973416	91973578	MACS_peak_406	50.81
chr2	91996073	91996238	MACS_peak_407	59.63
chr2	92243693	92243988	MACS_peak_408	68.48
chr2	92244148	92244441	MACS_peak_409	54.72
chr2	92268975	92270000	MACS_peak_410	239.11
chr2	92272870	92273125	MACS_peak_411	53.07
chr2	92276596	92276825	MACS_peak_412	64.53
chr2	92280454	92280671	MACS_peak_413	61.13
chr2	92281259	92281429	MACS_peak_414	186.33
chr2	92281446	92281544	MACS_peak_415	66.97
chr2	92289863	92291796	MACS_peak_416	416.45
chr2	92295715	92297028	MACS_peak_417	306.99
chr2	92297043	92298232	MACS_peak_418	530.58
chr2	92300506	92301351	MACS_peak_419	184.14
chr2	92305252	92305544	MACS_peak_420	116.31
chr2	92305623	92305825	MACS_peak_421	161.35
chr2	92305902	92307958	MACS_peak_422	491.07
chr2	92307966	92308409	MACS_peak_423	229.71
chr2	92308430	92309384	MACS_peak_424	460.7
chr2	92309440	92311206	MACS_peak_425	808.15

chr2	92311209	92314571	MACS_peak_426	1173.62
chr2	92315722	92322211	MACS_peak_427	2102.96
chr2	92322432	92325626	MACS_peak_428	1557.31
chr2	92325675	92326171	MACS_peak_429	208.07
chr2	109815780	109816312	MACS_peak_430	63.92
chr2	114231865	114232062	MACS_peak_431	54.33
chr2	128719874	128720007	MACS_peak_432	59.88
chr2	132975470	132975656	MACS_peak_433	53.14
chr2	133001970	133002152	MACS_peak_434	54.15
chr2	133003294	133003349	MACS_peak_435	55.67
chr2	134672534	134672665	MACS_peak_436	55.9
chr2	141287742	141287870	MACS_peak_437	61.1
chr2	142743854	142743994	MACS_peak_438	58.24
chr2	146039531	146039659	MACS_peak_439	53.13
chr2	158373847	158373932	MACS_peak_440	80.72
chr2	194773242	194773386	MACS_peak_441	54.02
chr2	216362227	216362375	MACS_peak_442	56.48
chr2	228176884	228177020	MACS_peak_443	59.16
chr2	232506384	232506479	MACS_peak_444	52.4
chr20	12148234	12148382	MACS_peak_445	61.35
chr20	20647633	20647788	MACS_peak_446	59.31
chr20	24360066	24360203	MACS_peak_447	58.93
chr20	26305690	26305823	MACS_peak_448	62.84
chr20	26314959	26315211	MACS_peak_449	57.45
chr20	29811927	29812017	MACS_peak_450	61
chr20	29815229	29815289	MACS_peak_451	52.84
chr20	29829299	29829783	MACS_peak_452	111.89
chr20	29832018	29832154	MACS_peak_453	52.45
chr20	36699488	36699639	MACS_peak_454	50.15
chr20	38588232	38588380	MACS_peak_455	50.3
chr20	38879547	38879677	MACS_peak_456	59.08
chr20	44142462	44142570	MACS_peak_457	54.59
chr20	48784734	48784868	MACS_peak_458	52.03
chr20	53912556	53912733	MACS_peak_459	54.83
chr20	55482589	55482726	MACS_peak_460	52.01
chr20	55707181	55707296	MACS_peak_461	63.43
chr20	56313167	56313310	MACS_peak_462	68.15
chr20	62947420	62947518	MACS_peak_463	63.79
chr21	9761290	9761441	MACS_peak_464	59.7
chr21	10126023	10126239	MACS_peak_465	50.71
chr21	10775138	10775440	MACS_peak_466	51.88
chr21	10819313	10819772	MACS_peak_467	161.59
chr21	10856728	10857109	MACS_peak_468	83.91

chr21	14353239	14353396	MACS_peak_469	78.17
chr21	35173737	35173881	MACS_peak_470	57.35
chr22	16861615	16861725	MACS_peak_471	72.59
chr22	16953263	16953398	MACS_peak_472	50.38
chr22	17352504	17352639	MACS_peak_473	61.28
chr22	17372331	17372465	MACS_peak_474	55.78
chr22	18879904	18880058	MACS_peak_475	56.38
chr22	51243685	51243845	MACS_peak_476	54.33
chr3	1687764	1687893	MACS_peak_477	58.31
chr3	3687190	3687331	MACS_peak_478	50.12
chr3	4258863	4258997	MACS_peak_479	55.19
chr3	8336101	8336219	MACS_peak_480	52.55
chr3	11914125	11914265	MACS_peak_481	57.76
chr3	25091328	25091468	MACS_peak_482	62.31
chr3	29815084	29815222	MACS_peak_483	51.79
chr3	30471812	30471986	MACS_peak_484	55.99
chr3	33552712	33552855	MACS_peak_485	54.27
chr3	36622968	36623112	MACS_peak_486	55.83
chr3	36662968	36663099	MACS_peak_487	59.32
chr3	40545128	40545258	MACS_peak_488	54.34
chr3	44650524	44650672	MACS_peak_489	53.58
chr3	58839192	58839318	MACS_peak_490	53.63
chr3	63420142	63420283	MACS_peak_491	58.01
chr3	65621469	65621606	MACS_peak_492	56.8
chr3	73911422	73911558	MACS_peak_493	53.63
chr3	80617709	80617837	MACS_peak_494	51.53
chr3	82132993	82133132	MACS_peak_495	54.22
chr3	86847573	86847728	MACS_peak_496	56.43
chr3	87503655	87503802	MACS_peak_497	65.01
chr3	89378680	89378818	MACS_peak_498	50.79
chr3	90394745	90394879	MACS_peak_499	53.88
chr3	90413914	90414099	MACS_peak_500	55.55
chr3	90417301	90417437	MACS_peak_501	54.33
chr3	90487246	90487347	MACS_peak_502	52.68
chr3	95045650	95045773	MACS_peak_503	55.05
chr3	100469364	100469499	MACS_peak_504	53.86
chr3	113208856	113209005	MACS_peak_505	52.23
chr3	117892395	117892520	MACS_peak_506	65.95
chr3	117933972	117934072	MACS_peak_507	64.54
chr3	122453948	122454051	MACS_peak_508	65.54
chr3	135037057	135037173	MACS_peak_509	70.2
chr3	135691890	135692014	MACS_peak_510	59.18
chr3	149533758	149533889	MACS_peak_511	52.41

chr3	157659971	157660110	MACS_peak_512	52.95
chr3	157697764	157697903	MACS_peak_513	58.47
chr3	178902750	178902896	MACS_peak_514	56.91
chr3	196625612	196625829	MACS_peak_515	286.91
chr3	197896033	197896237	MACS_peak_516	56.16
chr4	9357225	9357345	MACS_peak_517	55.16
chr4	11949755	11949873	MACS_peak_518	52.55
chr4	14785862	14786007	MACS_peak_519	54.61
chr4	15530377	15530497	MACS_peak_520	63.18
chr4	21583638	21583707	MACS_peak_521	58.79
chr4	31391367	31391515	MACS_peak_522	50.3
chr4	35967555	35967672	MACS_peak_523	56.97
chr4	37312637	37312743	MACS_peak_524	74.2
chr4	42583624	42583793	MACS_peak_525	59.2
chr4	49094741	49095769	MACS_peak_526	110.57
chr4	49096563	49097660	MACS_peak_527	120.74
chr4	49098162	49098378	MACS_peak_528	71.03
chr4	49098596	49099040	MACS_peak_529	130.81
chr4	49099058	49099416	MACS_peak_530	61.21
chr4	49099602	49099882	MACS_peak_531	62.9
chr4	49100162	49100359	MACS_peak_532	55.97
chr4	49100387	49102073	MACS_peak_533	188.92
chr4	49103117	49103236	MACS_peak_534	89.37
chr4	49103299	49105153	MACS_peak_535	137.96
chr4	49105155	49105750	MACS_peak_536	151.61
chr4	49105851	49106073	MACS_peak_537	111.99
chr4	49107997	49108330	MACS_peak_538	72.93
chr4	49109043	49111454	MACS_peak_539	87.65
chr4	49111495	49113124	MACS_peak_540	367.8
chr4	49119641	49119870	MACS_peak_541	80.74
chr4	49120302	49122341	MACS_peak_542	150.51
chr4	49122394	49123433	MACS_peak_543	155.3
chr4	49124660	49124979	MACS_peak_544	50.49
chr4	49125312	49125765	MACS_peak_545	61.16
chr4	49128513	49129206	MACS_peak_546	116.28
chr4	49130628	49131013	MACS_peak_547	51.97
chr4	49131941	49132228	MACS_peak_548	80.56
chr4	49133655	49134126	MACS_peak_549	64.23
chr4	49135386	49135517	MACS_peak_550	64.62
chr4	49136372	49136950	MACS_peak_551	82.45
chr4	49138432	49139405	MACS_peak_552	177.91
chr4	49139789	49140553	MACS_peak_553	51.59
chr4	49141647	49142140	MACS_peak_554	137.92

chr4	49142288	49143041	MACS_peak_555	108.44
chr4	49144089	49144411	MACS_peak_556	85.79
chr4	49146344	49146882	MACS_peak_557	109.45
chr4	49146993	49147203	MACS_peak_558	89.99
chr4	49147476	49148003	MACS_peak_559	59.43
chr4	49149240	49149925	MACS_peak_560	190.52
chr4	49150720	49151941	MACS_peak_561	231.71
chr4	49152389	49153340	MACS_peak_562	105.94
chr4	49154817	49155839	MACS_peak_563	136.07
chr4	49156030	49156899	MACS_peak_564	184.55
chr4	49157906	49158156	MACS_peak_565	71.37
chr4	49309597	49309809	MACS_peak_566	50.23
chr4	49512788	49512906	MACS_peak_567	50.54
chr4	49568942	49569058	MACS_peak_568	53.84
chr4	49633563	49634132	MACS_peak_569	50.97
chr4	49634509	49634820	MACS_peak_570	55.27
chr4	49635196	49635505	MACS_peak_571	58.65
chr4	49635912	49636625	MACS_peak_572	53.49
chr4	49636709	49636900	MACS_peak_573	58.02
chr4	49637133	49637304	MACS_peak_574	68.49
chr4	49637423	49637683	MACS_peak_575	92.28
chr4	49639047	49639251	MACS_peak_576	81.54
chr4	49639288	49639867	MACS_peak_577	69.32
chr4	49640052	49640202	MACS_peak_578	53.78
chr4	49641128	49641220	MACS_peak_579	58.6
chr4	49643783	49643905	MACS_peak_580	55.49
chr4	49644607	49645019	MACS_peak_581	59.24
chr4	49646613	49646891	MACS_peak_582	78.16
chr4	49649778	49650469	MACS_peak_583	75.24
chr4	49650635	49650862	MACS_peak_584	56.64
chr4	49652716	49652976	MACS_peak_585	94.11
chr4	49653376	49653727	MACS_peak_586	55.23
chr4	49658282	49658350	MACS_peak_587	57.08
chr4	49659530	49659898	MACS_peak_588	171.42
chr4	59837530	59837681	MACS_peak_589	55.55
chr4	68264194	68265105	MACS_peak_590	132.17
chr4	68265110	68265781	MACS_peak_591	128.72
chr4	72990104	72990203	MACS_peak_592	69.09
chr4	73284302	73284440	MACS_peak_593	58.7
chr4	76313224	76313363	MACS_peak_594	58.47
chr4	89760880	89761026	MACS_peak_595	56.91
chr4	129355864	129356000	MACS_peak_596	51.24
chr4	130970983	130971129	MACS_peak_597	64.61

chr4	179711119	179711259	MACS_peak_598	52.02
chr4	185972107	185972251	MACS_peak_599	55.63
chr5	9060200	9060348	MACS_peak_600	56.48
chr5	15275390	15275516	MACS_peak_601	50.21
chr5	21378134	21378268	MACS_peak_602	50.42
chr5	34231186	34231293	MACS_peak_603	55.62
chr5	35504273	35504409	MACS_peak_604	51.24
chr5	40346209	40346318	MACS_peak_605	59.32
chr5	41206511	41206657	MACS_peak_606	56.91
chr5	41302949	41303126	MACS_peak_607	60.07
chr5	42398650	42398779	MACS_peak_608	55.78
chr5	46380257	46380392	MACS_peak_609	54.76
chr5	49408949	49409281	MACS_peak_610	51.91
chr5	49409286	49409852	MACS_peak_611	61.79
chr5	49413394	49414517	MACS_peak_612	64.73
chr5	49415589	49416992	MACS_peak_613	59.8
chr5	49417912	49418539	MACS_peak_614	74.08
chr5	49420390	49421344	MACS_peak_615	55.37
chr5	49422222	49422868	MACS_peak_616	74.7
chr5	49424445	49425845	MACS_peak_617	64.68
chr5	49426614	49427886	MACS_peak_618	78.78
chr5	49428478	49429174	MACS_peak_619	62.84
chr5	49429212	49429750	MACS_peak_620	61.7
chr5	49429806	49430522	MACS_peak_621	96.27
chr5	49433186	49434200	MACS_peak_622	80.37
chr5	49434820	49435537	MACS_peak_623	102.46
chr5	49436344	49436889	MACS_peak_624	65.09
chr5	49436901	49437641	MACS_peak_625	61.46
chr5	49439117	49440505	MACS_peak_626	65.56
chr5	53585388	53585516	MACS_peak_627	58.56
chr5	54336653	54336803	MACS_peak_628	61.49
chr5	56921878	56922025	MACS_peak_629	51.2
chr5	57120615	57120746	MACS_peak_630	53.41
chr5	59014790	59014929	MACS_peak_631	55.94
chr5	63647986	63648125	MACS_peak_632	50.56
chr5	64903595	64903689	MACS_peak_633	61.64
chr5	65742167	65742315	MACS_peak_634	53.58
chr5	67704979	67705103	MACS_peak_635	54.8
chr5	69190969	69191120	MACS_peak_636	52.52
chr5	69931252	69931362	MACS_peak_637	52.49
chr5	71940919	71941067	MACS_peak_638	51.35
chr5	83091530	83091678	MACS_peak_639	52.44
chr5	90279780	90279936	MACS_peak_640	62.92

chr5	94343299	94343432	MACS_peak_641	59.32
chr5	95981958	95982085	MACS_peak_642	55.08
chr5	96762250	96762388	MACS_peak_643	50.79
chr5	96851198	96851350	MACS_peak_644	61.44
chr5	106865718	106865861	MACS_peak_645	56.26
chr5	113494889	113495023	MACS_peak_646	59.16
chr5	113821134	113821264	MACS_peak_647	54.34
chr5	117401818	117401941	MACS_peak_648	55.05
chr5	117709381	117709515	MACS_peak_649	59.73
chr5	122353066	122353165	MACS_peak_650	66.5
chr5	127940855	127941000	MACS_peak_651	50.76
chr5	131819638	131819768	MACS_peak_652	107.37
chr5	131822246	131822373	MACS_peak_653	72.51
chr5	131822643	131822824	MACS_peak_654	113.96
chr5	134090537	134090685	MACS_peak_655	52.07
chr5	135970062	135970206	MACS_peak_656	57.35
chr5	141547980	141548123	MACS_peak_657	57.57
chr5	147470541	147470679	MACS_peak_658	50.95
chr5	149845362	149845498	MACS_peak_659	73.4
chr5	162332291	162332421	MACS_peak_660	56.14
chr5	171223479	171223649	MACS_peak_661	53.23
chr5	174313540	174313653	MACS_peak_662	64.63
chr6	57372450	57372710	MACS_peak_663	83.53
chr6	58032376	58032525	MACS_peak_664	52.23
chr6	58773757	58773949	MACS_peak_665	124.36
chr6	58774079	58774207	MACS_peak_666	57.32
chr6	58774326	58774461	MACS_peak_667	117.7
chr6	58774490	58774819	MACS_peak_668	65.26
chr6	58776206	58780166	MACS_peak_669	3100
chr6	61914108	61914409	MACS_peak_670	102.17
chr6	86678987	86679127	MACS_peak_671	68.98
chr6	90600086	90600239	MACS_peak_672	65.78
chr6	132212869	132213008	MACS_peak_673	58.47
chr6	138345229	138345377	MACS_peak_674	56.48
chr6	142763124	142763239	MACS_peak_675	76.13
chr6	143089943	143090110	MACS_peak_676	62.66
chr6	152687957	152688100	MACS_peak_677	57.57
chr6	161060599	161060801	MACS_peak_678	66.02
chr6	169430023	169430169	MACS_peak_679	56.91
chr7	6614713	6614860	MACS_peak_680	54.79
chr7	12563700	12563846	MACS_peak_681	54.99
chr7	12673836	12673972	MACS_peak_682	56.63
chr7	25775604	25775702	MACS_peak_683	69.09

chr7	26975129	26975278	MACS_peak_684	67.65
chr7	27489389	27489535	MACS_peak_685	54.01
chr7	49155362	49155498	MACS_peak_686	55.85
chr7	56648996	56649109	MACS_peak_687	58.78
chr7	57658210	57658340	MACS_peak_688	55.78
chr7	57742179	57742302	MACS_peak_689	51.03
chr7	58043239	58043369	MACS_peak_690	61.71
chr7	61083360	61083786	MACS_peak_691	78.49
chr7	61152123	61152307	MACS_peak_692	64.04
chr7	61639346	61639539	MACS_peak_693	73.58
chr7	61739557	61739734	MACS_peak_694	53.43
chr7	61828168	61828286	MACS_peak_695	58.4
chr7	61840561	61840727	MACS_peak_696	54.59
chr7	61967156	61967750	MACS_peak_697	521.76
chr7	61967788	61968142	MACS_peak_698	501.51
chr7	61968201	61969752	MACS_peak_699	1944.13
chr7	61969817	61970630	MACS_peak_700	934.65
chr7	61970674	61971412	MACS_peak_701	117.55
chr7	61971622	61972499	MACS_peak_702	84.46
chr7	61972580	61973228	MACS_peak_703	344.63
chr7	61973962	61974076	MACS_peak_704	59.17
chr7	61974116	61974390	MACS_peak_705	80.04
chr7	61974648	61974746	MACS_peak_706	54.37
chr7	61974842	61975099	MACS_peak_707	120.45
chr7	61975365	61975584	MACS_peak_708	132.61
chr7	61976008	61976459	MACS_peak_709	205.71
chr7	61976475	61976895	MACS_peak_710	63.2
chr7	61976991	61977130	MACS_peak_711	97.55
chr7	61979244	61979447	MACS_peak_712	105.4
chr7	61980285	61980604	MACS_peak_713	59.26
chr7	61981818	61982261	MACS_peak_714	65.73
chr7	61988522	61988682	MACS_peak_715	84.6
chr7	62052654	62052786	MACS_peak_716	55.37
chr7	64963660	64963756	MACS_peak_717	52.24
chr7	75768406	75768530	MACS_peak_718	56.18
chr7	76065299	76065393	MACS_peak_719	51.69
chr7	78527114	78527251	MACS_peak_720	60.76
chr7	83619444	83619587	MACS_peak_721	52.6
chr7	85439925	85440034	MACS_peak_722	55.04
chr7	86375935	86376108	MACS_peak_723	52.44
chr7	94227540	94227696	MACS_peak_724	70.46
chr7	96891120	96891218	MACS_peak_725	67.32
chr7	103760974	103761124	MACS_peak_726	52.49

chr7	110327911	110328011	MACS_peak_727	64.92
chr7	113259291	113259444	MACS_peak_728	56.88
chr7	114808812	114808964	MACS_peak_729	57.11
chr7	115560242	115560387	MACS_peak_730	52.17
chr7	116831198	116831338	MACS_peak_731	52.9
chr7	119228476	119228632	MACS_peak_732	65.09
chr7	125016418	125016558	MACS_peak_733	51.17
chr7	140478652	140478782	MACS_peak_734	50.9
chr7	151896133	151896270	MACS_peak_735	60.76
chr7	151922033	151922284	MACS_peak_736	69.52
chr7	158743462	158743608	MACS_peak_737	50.72
chr8	31089068	31089226	MACS_peak_738	55.78
chr8	43094763	43094940	MACS_peak_739	83.33
chr8	43095071	43095333	MACS_peak_740	75.17
chr8	43095807	43096000	MACS_peak_741	79.57
chr8	43772255	43772537	MACS_peak_742	71.41
chr8	43793161	43794309	MACS_peak_743	317.13
chr8	43794657	43794880	MACS_peak_744	63.79
chr8	43798741	43798919	MACS_peak_745	67.18
chr8	43808558	43808703	MACS_peak_746	51.45
chr8	43818821	43818977	MACS_peak_747	52.45
chr8	43821290	43822017	MACS_peak_748	58.86
chr8	43822059	43822704	MACS_peak_749	150.63
chr8	43822712	43823667	MACS_peak_750	50.18
chr8	43824043	43824403	MACS_peak_751	133.07
chr8	43824444	43825500	MACS_peak_752	81.43
chr8	43825535	43826275	MACS_peak_753	82.81
chr8	43826316	43827592	MACS_peak_754	132.24
chr8	43827634	43828527	MACS_peak_755	139.95
chr8	43829742	43830209	MACS_peak_756	169.73
chr8	43830408	43832173	MACS_peak_757	99.75
chr8	43832178	43833492	MACS_peak_758	63.09
chr8	43833518	43834054	MACS_peak_759	135.07
chr8	43834094	43835632	MACS_peak_760	228.96
chr8	43835971	43836580	MACS_peak_761	74
chr8	43836616	43836924	MACS_peak_762	75.36
chr8	43836949	43837842	MACS_peak_763	176.9
chr8	43838334	43838847	MACS_peak_764	143.92
chr8	46838891	46839428	MACS_peak_765	72.88
chr8	46839747	46840834	MACS_peak_766	201.83
chr8	46841309	46841850	MACS_peak_767	104.15
chr8	46841908	46843315	MACS_peak_768	172.78
chr8	46843924	46845442	MACS_peak_769	366.07

chr8	46846701	46846978	MACS_peak_770	105.32
chr8	46848195	46848810	MACS_peak_771	54.87
chr8	46848899	46851060	MACS_peak_772	234.7
chr8	46851295	46851569	MACS_peak_773	77.47
chr8	46851629	46851854	MACS_peak_774	98.18
chr8	46852396	46854695	MACS_peak_775	509.46
chr8	46854777	46856452	MACS_peak_776	351.19
chr8	51146327	51146495	MACS_peak_777	65.5
chr8	53681806	53681974	MACS_peak_778	69.69
chr8	61478162	61478336	MACS_peak_779	54.33
chr8	73107459	73107581	MACS_peak_780	61.32
chr8	76656278	76656419	MACS_peak_781	50.61
chr8	82885369	82885516	MACS_peak_782	50.51
chr8	86795461	86795667	MACS_peak_783	56.95
chr8	87567151	87567287	MACS_peak_784	54.72
chr8	92741838	92742017	MACS_peak_785	62.59
chr8	109806072	109806227	MACS_peak_786	57.17
chr8	115364053	115364202	MACS_peak_787	51.86
chr8	116942042	116942194	MACS_peak_788	61.16
chr8	119562041	119562185	MACS_peak_789	54.83
chr8	141422295	141422435	MACS_peak_790	54.94
chr9	10234	10345	MACS_peak_791	142.15
chr9	1374839	1374969	MACS_peak_792	51.68
chr9	7929317	7929475	MACS_peak_793	60.73
chr9	9264610	9264730	MACS_peak_794	58.4
chr9	11459555	11459690	MACS_peak_795	53.12
chr9	13915142	13915275	MACS_peak_796	56.03
chr9	14797351	14797475	MACS_peak_797	55.31
chr9	20877194	20877309	MACS_peak_798	51.62
chr9	29543169	29543305	MACS_peak_799	52.75
chr9	34750024	34750173	MACS_peak_800	53.76
chr9	42819080	42819177	MACS_peak_801	83.15
chr9	44222968	44223098	MACS_peak_802	51.67
chr9	45486388	45486558	MACS_peak_803	72.38
chr9	46927581	46927721	MACS_peak_804	50.34
chr9	66504253	66504495	MACS_peak_805	59.57
chr9	66778907	66779016	MACS_peak_806	55.02
chr9	66780142	66780394	MACS_peak_807	73.58
chr9	66817254	66817538	MACS_peak_808	61.53
chr9	66820230	66820408	MACS_peak_809	77.51
chr9	66821691	66821750	MACS_peak_810	59.3
chr9	66833015	66833437	MACS_peak_811	70.2
chr9	66833746	66834068	MACS_peak_812	178.57

chr9	66962706	66962842	MACS_peak_813	54.72
chr9	66970957	66971510	MACS_peak_814	243.43
chr9	66973480	66973768	MACS_peak_815	79.32
chr9	66975962	66976145	MACS_peak_816	53.29
chr9	67072128	67072314	MACS_peak_817	62.42
chr9	67221373	67221493	MACS_peak_818	63.29
chr9	67318244	67318421	MACS_peak_819	88.72
chr9	67320244	67320984	MACS_peak_820	295.21
chr9	67351366	67351557	MACS_peak_821	54.52
chr9	69621400	69621536	MACS_peak_822	52.41
chr9	69690133	69690203	MACS_peak_823	78.21
chr9	69708573	69708885	MACS_peak_824	110.22
chr9	69710690	69711384	MACS_peak_825	273.7
chr9	69926639	69926768	MACS_peak_826	54.41
chr9	69950321	69950451	MACS_peak_827	70.57
chr9	69984663	69984725	MACS_peak_828	68
chr9	70009969	70010057	MACS_peak_829	62.73
chr9	70076647	70076803	MACS_peak_830	58.28
chr9	70206095	70206246	MACS_peak_831	63.41
chr9	70327443	70327628	MACS_peak_832	80.56
chr9	72652996	72653476	MACS_peak_833	67.4
chr9	72771337	72771492	MACS_peak_834	72.89
chr9	72923750	72923897	MACS_peak_835	54.33
chr9	73511129	73511268	MACS_peak_836	50.56
chr9	83822944	83823093	MACS_peak_837	54.15
chr9	89516696	89516883	MACS_peak_838	72.21
chr9	101987912	101988115	MACS_peak_839	75.37
chr9	102588676	102588792	MACS_peak_840	64.27
chr9	104057617	104057736	MACS_peak_841	58.95
chr9	106638289	106638406	MACS_peak_842	59.12
chr9	106804354	106804468	MACS_peak_843	58.85
chr9	118625426	118625565	MACS_peak_844	65.92
chr9	118630294	118630423	MACS_peak_845	54.69
chr9_gl000199_random	30859	31193	MACS_peak_846	67.48
chr9_gl000199_random	34305	34715	MACS_peak_847	165.97
chr9_gl000199_random	34978	35080	MACS_peak_848	56.15
chr9_gl000199_random	36757	37095	MACS_peak_849	50.77
chr9_gl000199_random	37211	37535	MACS_peak_850	109.35
chr9_gl000199_random	40024	41112	MACS_peak_851	93.93
chr9_gl000199_random	41145	41839	MACS_peak_852	227.95
chr9_gl000199_random	46107	47122	MACS_peak_853	68.88
chr9_gl000199_random	47155	48107	MACS_peak_854	57.97
chr9_gl000199_random	48234	49048	MACS_peak_855	64.15

chr9_gl000199_random	49402	50495	MACS_peak_856	173.5
chr9_gl000199_random	50707	52842	MACS_peak_857	154.2
chr9_gl000199_random	52941	53244	MACS_peak_858	95.89
chr9_gl000199_random	53250	53567	MACS_peak_859	58.6
chr9_gl000199_random	54352	55201	MACS_peak_860	144.08
chr9_gl000199_random	55351	56015	MACS_peak_861	84.14
chr9_gl000199_random	57054	57731	MACS_peak_862	219.12
chr9_gl000199_random	58169	58886	MACS_peak_863	71.42
chr9_gl000199_random	59309	60027	MACS_peak_864	143.78
chr9_gl000199_random	60533	64676	MACS_peak_865	64.26
chr9_gl000199_random	64742	67171	MACS_peak_866	138.8
chr9_gl000199_random	67199	68583	MACS_peak_867	56.79
chr9_gl000199_random	68618	68936	MACS_peak_868	79.14
chr9_gl000199_random	68985	69797	MACS_peak_869	124.1
chr9_gl000199_random	69968	70998	MACS_peak_870	68.83
chr9_gl000199_random	72067	73244	MACS_peak_871	182.93
chr9_gl000199_random	73735	74335	MACS_peak_872	56.9
chr9_gl000199_random	74351	74892	MACS_peak_873	81.29
chr9_gl000199_random	74907	78439	MACS_peak_874	83.08
chr9_gl000199_random	78477	81322	MACS_peak_875	299.19
chr9_gl000199_random	82762	83589	MACS_peak_876	155.06
chr9_gl000199_random	86137	86602	MACS_peak_877	72.18
chr9_gl000199_random	87695	88303	MACS_peak_878	84.12
chr9_gl000199_random	88877	89932	MACS_peak_879	111.39
chr9_gl000199_random	90737	94806	MACS_peak_880	362.15
chr9_gl000199_random	94862	96941	MACS_peak_881	166.74
chr9_gl000199_random	98621	100016	MACS_peak_882	103.26
chr9_gl000199_random	101358	103538	MACS_peak_883	130.67
chr9_gl000199_random	105030	107504	MACS_peak_884	257.63
chr9_gl000199_random	107547	110024	MACS_peak_885	186.57
chr9_gl000199_random	110028	110947	MACS_peak_886	102.97
chr9_gl000199_random	113431	115697	MACS_peak_887	248.17
chr9_gl000199_random	115734	116903	MACS_peak_888	105.69
chr9_gl000199_random	116905	117629	MACS_peak_889	67.42
chr9_gl000199_random	118197	123384	MACS_peak_890	262.07
chr9_gl000199_random	123387	125070	MACS_peak_891	230.01
chr9_gl000199_random	125071	125329	MACS_peak_892	79.42
chr9_gl000199_random	129568	130166	MACS_peak_893	111.21
chr9_gl000199_random	131164	131327	MACS_peak_894	55.77
chr9_gl000199_random	133365	134085	MACS_peak_895	90.88
chr9_gl000199_random	134087	134592	MACS_peak_896	54.45
chr9_gl000199_random	134595	136301	MACS_peak_897	90.95
chr9_gl000199_random	136686	137086	MACS_peak_898	55.03

chr9_gl000199_random	137138	137557	MACS_peak_899	82.12
chr9_gl000199_random	137748	138596	MACS_peak_900	98.53
chr9_gl000199_random	139002	139309	MACS_peak_901	53.1
chr9_gl000199_random	140083	141757	MACS_peak_902	123.05
chr9_gl000199_random	142281	143128	MACS_peak_903	70.1
chr9_gl000199_random	144599	145148	MACS_peak_904	62.78
chr9_gl000199_random	145192	147644	MACS_peak_905	240.89
chr9_gl000199_random	147921	148823	MACS_peak_906	131.86
chr9_gl000199_random	149072	150713	MACS_peak_907	109.09
chr9_gl000199_random	150759	152111	MACS_peak_908	117.57
chr9_gl000199_random	152289	157215	MACS_peak_909	211.32
chr9_gl000199_random	158970	159476	MACS_peak_910	119.1
chr9_gl000199_random	159542	161102	MACS_peak_911	110.07
chr9_gl000199_random	161675	162447	MACS_peak_912	69.15
chr9_gl000199_random	162692	164230	MACS_peak_913	90.67
chr9_gl000199_random	164243	164876	MACS_peak_914	76.54
chrUn_gl000212	40722	41019	MACS_peak_915	73.94
chrUn_gl000214	64314	65752	MACS_peak_916	188.92
chrUn_gl000214	120216	121588	MACS_peak_917	52.23
chrUn_gl000216	1183	1674	MACS_peak_918	73.53
chrUn_gl000216	1727	1933	MACS_peak_919	88.4
chrUn_gl000216	4010	4469	MACS_peak_920	50.77
chrUn_gl000216	7325	7569	MACS_peak_921	63.79
chrUn_gl000216	9240	9794	MACS_peak_922	111.56
chrUn_gl000216	11201	11481	MACS_peak_923	65.1
chrUn_gl000216	13912	14006	MACS_peak_924	69.17
chrUn_gl000216	14901	15286	MACS_peak_925	66.89
chrUn_gl000216	18191	18686	MACS_peak_926	76.65
chrUn_gl000216	22422	22702	MACS_peak_927	63.16
chrUn_gl000216	31107	31197	MACS_peak_928	57.56
chrUn_gl000216	75575	76189	MACS_peak_929	137.67
chrUn_gl000216	80615	80884	MACS_peak_930	92.87
chrUn_gl000216	144954	145614	MACS_peak_931	113.83
chrUn_gl000216	148762	148936	MACS_peak_932	77.51
chrUn_gl000216	151130	151480	MACS_peak_933	72.98
chrUn_gl000216	151678	152034	MACS_peak_934	60.84
chrUn_gl000216	154276	154501	MACS_peak_935	110
chrUn_gl000216	154628	154794	MACS_peak_936	72.87
chrUn_gl000216	155225	155398	MACS_peak_937	112.68
chrUn_gl000216	159968	160383	MACS_peak_938	97.26
chrUn_gl000216	160465	160715	MACS_peak_939	91.89
chrUn_gl000216	161589	161687	MACS_peak_940	92.21
chrUn_gl000216	163128	163316	MACS_peak_941	54.64

chrUn_gl000216	163427	163772	MACS_peak_942	196.73
chrUn_gl000216	165840	166060	MACS_peak_943	64.58
chrUn_gl000217	139776	140066	MACS_peak_944	63
chrUn_gl000218	131086	131178	MACS_peak_945	51.53
chrUn_gl000219	3797	3895	MACS_peak_946	56
chrUn_gl000221	108433	108682	MACS_peak_947	60.32
chrUn_gl000224	6679	9812	MACS_peak_948	101.44
chrUn_gl000224	15136	15597	MACS_peak_949	53.6
chrUn_gl000224	17755	19041	MACS_peak_950	54.73
chrUn_gl000224	160777	161027	MACS_peak_951	55.06
chrUn_gl000225	6154	6323	MACS_peak_952	53.61
chrUn_gl000225	6596	6979	MACS_peak_953	118.07
chrUn_gl000225	13888	14150	MACS_peak_954	68.36
chrUn_gl000225	14369	14597	MACS_peak_955	60.36
chrUn_gl000225	21501	21972	MACS_peak_956	81.8
chrUn_gl000225	24744	25314	MACS_peak_957	64.08
chrUn_gl000225	25429	25799	MACS_peak_958	143.26
chrUn_gl000225	33214	33465	MACS_peak_959	79.02
chrUn_gl000225	33531	33830	MACS_peak_960	103.01
chrUn_gl000225	45911	46207	MACS_peak_961	58.61
chrUn_gl000225	49574	49853	MACS_peak_962	91.79
chrUn_gl000225	49933	50258	MACS_peak_963	53.01
chrUn_gl000225	50762	51008	MACS_peak_964	168.91
chrUn_gl000225	56053	56285	MACS_peak_965	106.95
chrUn_gl000225	56526	56747	MACS_peak_966	68.42
chrUn_gl000225	61411	61940	MACS_peak_967	62.78
chrUn_gl000225	67153	67599	MACS_peak_968	200.02
chrUn_gl000225	74414	74922	MACS_peak_969	161.35
chrUn_gl000225	75483	75636	MACS_peak_970	59.82
chrUn_gl000225	85545	85735	MACS_peak_971	172.77
chrUn_gl000225	86449	86694	MACS_peak_972	52.86
chrUn_gl000225	88018	88352	MACS_peak_973	55.85
chrUn_gl000225	92230	92567	MACS_peak_974	66.99
chrUn_gl000225	96987	97243	MACS_peak_975	81.59
chrUn_gl000225	101536	101765	MACS_peak_976	83.9
chrUn_gl000225	102069	102263	MACS_peak_977	52.37
chrUn_gl000225	103271	103364	MACS_peak_978	58.64
chrUn_gl000225	107143	107558	MACS_peak_979	58.55
chrUn_gl000225	110203	110646	MACS_peak_980	115.79
chrUn_gl000225	111036	111350	MACS_peak_981	65.57
chrUn_gl000225	111446	112020	MACS_peak_982	103.38
chrUn_gl000225	116150	116286	MACS_peak_983	51.49
chrUn_gl000225	116495	116610	MACS_peak_984	77.38

chrUn_gl000225	116630	117120	MACS_peak_985	169.85
chrUn_gl000225	117150	117229	MACS_peak_986	57.23
chrUn_gl000225	117721	118027	MACS_peak_987	133.08
chrUn_gl000225	118283	118579	MACS_peak_988	54.08
chrUn_gl000225	120067	120237	MACS_peak_989	53.91
chrUn_gl000225	120437	120793	MACS_peak_990	124.56
chrUn_gl000225	125029	125419	MACS_peak_991	61.23
chrUn_gl000225	127177	127394	MACS_peak_992	52.61
chrUn_gl000225	129069	129351	MACS_peak_993	68.54
chrUn_gl000225	130376	130630	MACS_peak_994	77.74
chrUn_gl000225	131508	131844	MACS_peak_995	138.9
chrUn_gl000225	133609	133929	MACS_peak_996	78.08
chrUn_gl000226	0	297	MACS_peak_997	231.11
chrUn_gl000228	116420	117372	MACS_peak_998	86.55
chrUn_gl000228	117602	117881	MACS_peak_999	87.42
chrUn_gl000228	117907	118946	MACS_peak_1000	387.15
chrUn_gl000237	33956	34143	MACS_peak_1001	50.62
chrX	48085923	48086049	MACS_peak_1002	59.32
chrX	58549283	58549357	MACS_peak_1003	62.62
chrX	58561374	58561545	MACS_peak_1004	130.77
chrX	58561562	58561720	MACS_peak_1005	51.93
chrX	58564281	58564606	MACS_peak_1006	89.38
chrX	58564720	58582010	MACS_peak_1007	3185.05
chrX	61682078	61683924	MACS_peak_1008	239.63
chrX	61683971	61684509	MACS_peak_1009	91.22
chrX	61684594	61684857	MACS_peak_1010	74.04
chrX	61684903	61685546	MACS_peak_1011	182.16
chrX	61686490	61687639	MACS_peak_1012	140.06
chrX	61688707	61689723	MACS_peak_1013	151.48
chrX	61689763	61690426	MACS_peak_1014	218.66
chrX	61690771	61693090	MACS_peak_1015	92.61
chrX	61693849	61694821	MACS_peak_1016	169.99
chrX	61694882	61696288	MACS_peak_1017	139.15
chrX	61696359	61697052	MACS_peak_1018	131.54
chrX	61697622	61698807	MACS_peak_1019	124.9
chrX	61700305	61700877	MACS_peak_1020	152.83
chrX	61700917	61701166	MACS_peak_1021	78.32
chrX	61701228	61701667	MACS_peak_1022	56.83
chrX	61701736	61702640	MACS_peak_1023	76.05
chrX	61702695	61703139	MACS_peak_1024	82.33
chrX	61703525	61707160	MACS_peak_1025	347.32
chrX	61707620	61707843	MACS_peak_1026	91.18
chrX	61708713	61710847	MACS_peak_1027	318.47

chrX	61711181	61711897	MACS_peak_1028	192.36
chrX	61711917	61712866	MACS_peak_1029	187.2
chrX	61713151	61714303	MACS_peak_1030	93.03
chrX	61714371	61715610	MACS_peak_1031	140.63
chrX	61715639	61716931	MACS_peak_1032	92.87
chrX	61717720	61718279	MACS_peak_1033	157.38
chrX	61718597	61720149	MACS_peak_1034	272.61
chrX	61720495	61721063	MACS_peak_1035	223.72
chrX	61721158	61721579	MACS_peak_1036	57.65
chrX	61722617	61723243	MACS_peak_1037	133.99
chrX	61729215	61729963	MACS_peak_1038	72.57
chrX	61731020	61731262	MACS_peak_1039	59.86
chrX	61733738	61734194	MACS_peak_1040	51.17
chrX	61747264	61747326	MACS_peak_1041	55.06
chrX	61792707	61792776	MACS_peak_1042	51.23
chrX	61839213	61839407	MACS_peak_1043	104.01
chrX	61871560	61871751	MACS_peak_1044	52.95
chrX	69735981	69736118	MACS_peak_1045	54.49
chrX	73371005	73371148	MACS_peak_1046	57.57
chrX	89904459	89904561	MACS_peak_1047	68.45
chrX	93739326	93739446	MACS_peak_1048	63.18
chrX	110573972	110574120	MACS_peak_1049	56.48
chrX	130885838	130885978	MACS_peak_1050	58.24
chrX	134855276	134855531	MACS_peak_1051	54.3
chrY	9919180	9919302	MACS_peak_1052	52.07
chrY	9958892	9958992	MACS_peak_1053	59.32
chrY	13137785	13137870	MACS_peak_1054	68.11
chrY	13262134	13262260	MACS_peak_1055	53.63
chrY	13448625	13448855	MACS_peak_1056	61.27
chrY	13453091	13453529	MACS_peak_1057	105.05
chrY	13455503	13456090	MACS_peak_1058	73.53
chrY	13458355	13458501	MACS_peak_1059	58.86
chrY	13459700	13459773	MACS_peak_1060	62.53
chrY	13460947	13461422	MACS_peak_1061	97.02
chrY	13467836	13468392	MACS_peak_1062	186.46
chrY	13468741	13468982	MACS_peak_1063	51.27
chrY	13468993	13469208	MACS_peak_1064	72.1
chrY	13482022	13482187	MACS_peak_1065	148.34
chrY	13482232	13482424	MACS_peak_1066	73.28
chrY	13485517	13485799	MACS_peak_1067	149.97
chrY	13485855	13486002	MACS_peak_1068	72.1
chrY	13597850	13597981	MACS_peak_1069	60.36
chrY	13638418	13638615	MACS_peak_1070	69.79

chrY	13663815	13663894	MACS_peak_1071	50.33
chrY	13669475	13669589	MACS_peak_1072	77.53
chrY	13680933	13681014	MACS_peak_1073	69.83
chrY	13711073	13711200	MACS_peak_1074	58.08
chrY	13801063	13801196	MACS_peak_1075	99.19
chrY	13809828	13809898	MACS_peak_1076	50.81
chrY	13811036	13811330	MACS_peak_1077	83.84
chrY	13822607	13822667	MACS_peak_1078	74.6
chrY	13831675	13831733	MACS_peak_1079	62.9
chrY	13842627	13842842	MACS_peak_1080	239.31
chrY	13846398	13846528	MACS_peak_1081	84.66
chrY	13851943	13852100	MACS_peak_1082	75.35
chrY	13864451	13864656	MACS_peak_1083	69.81
chrY	13868367	13868584	MACS_peak_1084	126.41
chrY	13869033	13869269	MACS_peak_1085	58.59
chrY	58826810	58826903	MACS_peak_1086	56.35
chrY	58837517	58837590	MACS_peak_1087	68.61
chrY	58997238	58997403	MACS_peak_1088	63.67

Appendix III: ZNF350 biological replicate 2 peaks generated by ChIP-seq

Chromo	*Start*	*End*	*Peak Identifier*	*Peak Size*
chr1	245013	245146	MACS_peak_1	52.93
chr1	16885790	16886385	MACS_peak_2	50.83
chr1	16942160	16942629	MACS_peak_3	58.87
chr1	19171739	19171890	MACS_peak_4	51.04
chr1	19326379	19326627	MACS_peak_5	118.93
chr1	22733331	22733483	MACS_peak_6	50.8
chr1	31065713	31065873	MACS_peak_7	65.54
chr1	37012341	37012507	MACS_peak_8	100.43
chr1	39736006	39736209	MACS_peak_9	54
chr1	40013902	40014057	MACS_peak_10	50.08
chr1	51822427	51822577	MACS_peak_11	58.89
chr1	55541246	55541357	MACS_peak_12	69.95
chr1	56444650	56444813	MACS_peak_13	53.54
chr1	56774211	56774363	MACS_peak_14	58.34
chr1	61458024	61458163	MACS_peak_15	54.14
chr1	76604340	76604490	MACS_peak_16	51.29
chr1	77663529	77663664	MACS_peak_17	51.67
chr1	78353540	78353697	MACS_peak_18	55.62
chr1	100992503	100992658	MACS_peak_19	102.75
chr1	102123371	102123579	MACS_peak_20	120.84
chr1	107369914	107370107	MACS_peak_21	81.15
chr1	110687399	110687554	MACS_peak_22	50.08
chr1	112784076	112784211	MACS_peak_23	51.67
chr1	116304061	116304188	MACS_peak_24	224.13
chr1	120288222	120288376	MACS_peak_25	50.31
chr1	145167946	145168174	MACS_peak_26	51.73
chr1	153741052	153741199	MACS_peak_27	52.04
chr1	156186370	156186520	MACS_peak_28	61.18
chr1	160806311	160806468	MACS_peak_29	58.99
chr1	161736691	161736849	MACS_peak_30	51.27
chr1	186387912	186388041	MACS_peak_31	56.2
chr1	186924614	186924772	MACS_peak_32	54.13
chr1	200837383	200837513	MACS_peak_33	54.47
chr1	207580657	207580892	MACS_peak_34	58.79
chr1	208326919	208327067	MACS_peak_35	57.86
chr1	209889073	209889273	MACS_peak_36	58.1
chr1	228716246	228716429	MACS_peak_37	63.51
chr1	233474885	233475093	MACS_peak_38	54.05
chr1	240847659	240847827	MACS_peak_39	59.33

chr1	246804814	246805063	MACS_peak_40	54.37
chr1	247605679	247605821	MACS_peak_41	54.48
chr10	2126486	2126718	MACS_peak_42	97.06
chr10	7608760	7608941	MACS_peak_43	85.88
chr10	8723411	8723558	MACS_peak_44	75.09
chr10	14867975	14868122	MACS_peak_45	69.41
chr10	23639713	23639892	MACS_peak_46	107.28
chr10	26042270	26042487	MACS_peak_47	70.35
chr10	32110761	32110913	MACS_peak_48	50.8
chr10	32404196	32404378	MACS_peak_49	84.44
chr10	32404430	32404634	MACS_peak_50	108.36
chr10	49995844	49996101	MACS_peak_51	89.19
chr10	50394024	50394204	MACS_peak_52	52.26
chr10	71299190	71299327	MACS_peak_53	54.68
chr10	84802046	84802213	MACS_peak_54	55.29
chr10	84991570	84991723	MACS_peak_55	50.56
chr10	88163333	88163397	MACS_peak_56	388.37
chr10	94490333	94490435	MACS_peak_57	75.25
chr10	119174663	119174779	MACS_peak_58	56.7
chr10	124622630	124622792	MACS_peak_59	56.53
chr10	127297142	127297319	MACS_peak_60	61
chr11	3943026	3943172	MACS_peak_61	52.3
chr11	23836217	23836324	MACS_peak_62	94.83
chr11	45636835	45637004	MACS_peak_63	63.07
chr11	45815772	45815923	MACS_peak_64	52
chr11	47127259	47127414	MACS_peak_65	50.08
chr11	48914914	48915080	MACS_peak_66	55.62
chr11	50241869	50242063	MACS_peak_67	56.95
chr11	54915370	54915605	MACS_peak_68	53.84
chr11	57170988	57171122	MACS_peak_69	107.28
chr11	61214596	61214743	MACS_peak_70	52.04
chr11	64887648	64887791	MACS_peak_71	53.07
chr11	71844550	71844754	MACS_peak_72	68.97
chr11	73262241	73262363	MACS_peak_73	55.47
chr11	89992074	89992261	MACS_peak_74	72.67
chr11	97662327	97662491	MACS_peak_75	72.2
chr11	100932483	100932611	MACS_peak_76	57.26
chr11	101007733	101007881	MACS_peak_77	51.79
chr11	112580047	112580210	MACS_peak_78	83.65
chr11	126793061	126793218	MACS_peak_79	71.71
chr11	129960619	129960773	MACS_peak_80	50.31
chr11	133010818	133011003	MACS_peak_81	103.09
chr11	133224101	133224250	MACS_peak_82	51.54

chr12	5041729	5041840	MACS_peak_83	82.51
chr12	13055585	13055720	MACS_peak_84	50.35
chr12	14949155	14949293	MACS_peak_85	58.1
chr12	17745808	17745963	MACS_peak_86	50.08
chr12	54091939	54092057	MACS_peak_87	56.74
chr12	67266986	67267119	MACS_peak_88	50.25
chr12	69218280	69218430	MACS_peak_89	55.62
chr12	69222556	69222719	MACS_peak_90	72.83
chr12	69229618	69229769	MACS_peak_91	60.72
chr12	69233060	69233631	MACS_peak_92	392.29
chr12	95244987	95245128	MACS_peak_93	53.6
chr12	95825652	95825862	MACS_peak_94	82.17
chr12	95900929	95901065	MACS_peak_95	53.85
chr12	96192653	96192782	MACS_peak_96	50.13
chr12	100649904	100650043	MACS_peak_97	52.67
chr12	104415564	104415855	MACS_peak_98	119.71
chr12	106199277	106199415	MACS_peak_99	52.23
chr12	108350895	108351021	MACS_peak_100	56.09
chr12	110973480	110973659	MACS_peak_101	58.77
chr12	113150685	113150909	MACS_peak_102	149.06
chr12	126496534	126496630	MACS_peak_103	129.88
chr12	131653713	131653866	MACS_peak_104	50.56
chr13	21816958	21817147	MACS_peak_105	88.98
chr13	26257714	26257855	MACS_peak_106	50.06
chr13	29161486	29161591	MACS_peak_107	67.03
chr13	30759840	30759981	MACS_peak_108	53.6
chr13	34073656	34073802	MACS_peak_109	66.78
chr13	36982457	36982576	MACS_peak_110	58.55
chr13	51117179	51117334	MACS_peak_111	54.45
chr13	56274892	56275065	MACS_peak_112	76.27
chr13	56885020	56885173	MACS_peak_113	50.56
chr13	60269243	60269392	MACS_peak_114	50.09
chr13	64090471	64090578	MACS_peak_115	64.16
chr13	68604554	68604672	MACS_peak_116	58.87
chr13	70532644	70532810	MACS_peak_117	50.97
chr13	91830372	91830545	MACS_peak_118	55.47
chr13	96529946	96530113	MACS_peak_119	52.07
chr13	98838494	98838647	MACS_peak_120	58.88
chr13	102813928	102814076	MACS_peak_121	78.86
chr13	108706923	108707095	MACS_peak_122	55.1
chr13	109889257	109889421	MACS_peak_123	52.52
chr13	112859800	112859949	MACS_peak_124	51.54
chr14	19984298	19984455	MACS_peak_125	51.54

chr14	20782029	20782172	MACS_peak_126	53.07
chr14	23867124	23867227	MACS_peak_127	65.27
chr14	44474426	44474578	MACS_peak_128	59.15
chr14	64012777	64012904	MACS_peak_129	90.93
chr14	73347919	73348096	MACS_peak_130	125.84
chr14	80388080	80388230	MACS_peak_131	55.62
chr14	84796416	84796581	MACS_peak_132	72.86
chr15	20032570	20032740	MACS_peak_133	61.82
chr15	38973423	38973613	MACS_peak_134	57
chr15	44195941	44196085	MACS_peak_135	52.81
chr15	51202636	51202797	MACS_peak_136	56.79
chr15	58029241	58029406	MACS_peak_137	64.15
chr15	79248577	79248656	MACS_peak_138	107.67
chr15	79605741	79605893	MACS_peak_139	50.8
chr15	83389305	83389418	MACS_peak_140	572.21
chr15	84492647	84492735	MACS_peak_141	80.67
chr15	85852103	85852202	MACS_peak_142	56.68
chr15	91591419	91591580	MACS_peak_143	91.08
chr15	97097948	97098115	MACS_peak_144	51.45
chr16	20729219	20729424	MACS_peak_145	78.8
chr16	22674297	22674483	MACS_peak_146	82.13
chr16	22971036	22971129	MACS_peak_147	255.85
chr16	24468164	24468315	MACS_peak_148	51.04
chr16	28063422	28063577	MACS_peak_149	50.08
chr16	29290278	29290421	MACS_peak_150	53.07
chr16	30930153	30930312	MACS_peak_151	55.62
chr16	32003219	32003362	MACS_peak_152	53.07
chr16	33547490	33547887	MACS_peak_153	67.96
chr16	50520071	50520205	MACS_peak_154	54.85
chr16	51081041	51081131	MACS_peak_155	90.86
chr16	52058810	52058989	MACS_peak_156	84.83
chr16	54473990	54474247	MACS_peak_157	93.66
chr16	58436308	58436464	MACS_peak_158	54.94
chr16	64955818	64955939	MACS_peak_159	59.03
chr16	71866053	71866233	MACS_peak_160	103.88
chr16	72070018	72070170	MACS_peak_161	50.8
chr16	72831278	72831434	MACS_peak_162	55.62
chr16	76364039	76364180	MACS_peak_163	54.76
chr16	80426019	80426169	MACS_peak_164	51.29
chr16	82301218	82301338	MACS_peak_165	57.41
chr17	7572810	7572964	MACS_peak_166	50.31
chr17	7573948	7574092	MACS_peak_167	52.81
chr17	7578403	7578547	MACS_peak_168	52.81

chr17	39314731	39314853	MACS_peak_169	63.51
chr17	47617935	47618090	MACS_peak_170	50.08
chr17	54955382	54955566	MACS_peak_171	79.95
chr17	58998229	58998286	MACS_peak_172	124.08
chr17	60165619	60165828	MACS_peak_173	465.78
chr17	61994395	61994544	MACS_peak_174	59.98
chr17	64175985	64176179	MACS_peak_175	52.81
chr17	67366863	67366982	MACS_peak_176	66.19
chr17	68265226	68265381	MACS_peak_177	50.08
chr17	70142924	70143179	MACS_peak_178	95.01
chr17	71890740	71890878	MACS_peak_179	54.41
chr17	75635867	75635960	MACS_peak_180	85.61
chr17	80254113	80254321	MACS_peak_181	61.63
chr17_ctg5_hap1	1209611	1209721	MACS_peak_182	138.2
chr18	19648911	19649038	MACS_peak_183	76.18
chr18	31216202	31216365	MACS_peak_184	73.47
chr18	45401088	45401243	MACS_peak_185	50.08
chr18	55189223	55189370	MACS_peak_186	52.04
chr18	66235033	66235140	MACS_peak_187	64.16
chr19	7426938	7427095	MACS_peak_188	124.3
chr19	16448182	16448325	MACS_peak_189	53.07
chr19	18320530	18320643	MACS_peak_190	62.04
chr19	33604526	33604686	MACS_peak_191	62.09
chr19	34109278	34109417	MACS_peak_192	62.88
chr19	34697809	34697949	MACS_peak_193	53.5
chr19	39762996	39763143	MACS_peak_194	52.04
chr19	45085837	45085926	MACS_peak_195	107.67
chr19	52468027	52469506	MACS_peak_196	2494.81
chr19	52471831	52471993	MACS_peak_197	82.88
chr19	52472236	52472391	MACS_peak_198	114.84
chr19	57522482	57522594	MACS_peak_199	103.16
chr2	11094053	11094186	MACS_peak_200	52.23
chr2	19619591	19619740	MACS_peak_201	52.53
chr2	24290260	24290443	MACS_peak_202	55.34
chr2	26811837	26812055	MACS_peak_203	103.88
chr2	30924977	30925123	MACS_peak_204	50.84
chr2	37279096	37279212	MACS_peak_205	51.1
chr2	37420715	37420880	MACS_peak_206	245.75
chr2	48791546	48791720	MACS_peak_207	96.3
chr2	49354887	49355087	MACS_peak_208	92.53
chr2	62938669	62938812	MACS_peak_209	53.07
chr2	64429994	64430146	MACS_peak_210	50.8
chr2	67112337	67112487	MACS_peak_211	51.29

chr2	68743693	68743852	MACS_peak_212	57.3
chr2	81549575	81549767	MACS_peak_213	57.6
chr2	89629909	89630041	MACS_peak_214	74.9
chr2	95490892	95491068	MACS_peak_215	78.44
chr2	108301216	108301368	MACS_peak_216	59.15
chr2	113569664	113569876	MACS_peak_217	63.8
chr2	113775122	113775270	MACS_peak_218	78.28
chr2	119330202	119330336	MACS_peak_219	135.68
chr2	135874977	135875255	MACS_peak_220	112.91
chr2	137647861	137647942	MACS_peak_221	85.61
chr2	142301811	142302001	MACS_peak_222	60.09
chr2	143312853	143313003	MACS_peak_223	51.29
chr2	183078544	183078679	MACS_peak_224	64.1
chr2	193767548	193767700	MACS_peak_225	50.8
chr2	198761753	198761928	MACS_peak_226	98
chr2	221410880	221411224	MACS_peak_227	326.27
chr2	222431432	222431669	MACS_peak_228	95.62
chr2	224490820	224490913	MACS_peak_229	64.05
chr2	229629315	229629476	MACS_peak_230	83.21
chr20	1396641	1396938	MACS_peak_231	121.89
chr20	2662824	2663041	MACS_peak_232	78.36
chr20	4099249	4099423	MACS_peak_233	87.71
chr20	13305346	13305491	MACS_peak_234	50.39
chr20	22519381	22519528	MACS_peak_235	97.96
chr20	22519544	22519700	MACS_peak_236	167.41
chr20	26105987	26106192	MACS_peak_237	52.46
chr20	30528295	30528428	MACS_peak_238	55.8
chr20	32211649	32211804	MACS_peak_239	50.08
chr20	32620124	32620275	MACS_peak_240	59.43
chr20	35666647	35666817	MACS_peak_241	51.31
chr20	38335709	38335861	MACS_peak_242	50.8
chr20	43485492	43485624	MACS_peak_243	62.05
chr20	44832891	44833035	MACS_peak_244	52.81
chr20	44978191	44978352	MACS_peak_245	56.79
chr20	46596111	46596269	MACS_peak_246	50.85
chr20	47481884	47482035	MACS_peak_247	52
chr20	49623059	49623224	MACS_peak_248	55.78
chr20	50853812	50853969	MACS_peak_249	50.44
chr20	52820931	52821087	MACS_peak_250	50.03
chr20	52826789	52827057	MACS_peak_251	61.52
chr21	9682551	9682714	MACS_peak_252	52.46
chr21	37653787	37653935	MACS_peak_253	51.79
chr22	32820619	32820843	MACS_peak_254	127.62

chr22	33485231	33485382	MACS_peak_255	51.04
chr22	49392257	49392415	MACS_peak_256	55.62
chr3	2329250	2329474	MACS_peak_257	93.04
chr3	3326488	3326669	MACS_peak_258	55.52
chr3	13898392	13898595	MACS_peak_259	138.53
chr3	14762038	14762111	MACS_peak_260	118.47
chr3	29858421	29858556	MACS_peak_261	51.33
chr3	34172885	34173044	MACS_peak_262	68.35
chr3	40621055	40621211	MACS_peak_263	64.06
chr3	43171661	43171806	MACS_peak_264	52.55
chr3	46224046	46224212	MACS_peak_265	156.11
chr3	49333287	49333430	MACS_peak_266	53.07
chr3	51363964	51364130	MACS_peak_267	55.08
chr3	52815229	52815375	MACS_peak_268	52.3
chr3	55754141	55754278	MACS_peak_269	191.88
chr3	58655409	58655532	MACS_peak_270	116.78
chr3	63336900	63337048	MACS_peak_271	50.7
chr3	63749835	63749968	MACS_peak_272	63.51
chr3	65854995	65855179	MACS_peak_273	62.94
chr3	73178769	73178846	MACS_peak_274	82.89
chr3	84958156	84958263	MACS_peak_275	63.4
chr3	87106954	87107096	MACS_peak_276	53.34
chr3	93505768	93505921	MACS_peak_277	60.42
chr3	93517497	93517715	MACS_peak_278	55.52
chr3	113802032	113802199	MACS_peak_279	200.22
chr3	117220818	117221032	MACS_peak_280	59.35
chr3	119340979	119341122	MACS_peak_281	61.69
chr3	119891575	119891730	MACS_peak_282	63.51
chr3	121532952	121533118	MACS_peak_283	55.8
chr3	131519135	131519307	MACS_peak_284	79.67
chr3	132824580	132824727	MACS_peak_285	50.95
chr3	141197681	141197849	MACS_peak_286	167.85
chr3	142306806	142306975	MACS_peak_287	55.62
chr3	156590914	156591057	MACS_peak_288	52.22
chr3	156996251	156996387	MACS_peak_289	54.96
chr3	159381020	159381171	MACS_peak_290	50.32
chr3	184335115	184335255	MACS_peak_291	63.51
chr3	193019493	193019627	MACS_peak_292	55.52
chr4	11456549	11456663	MACS_peak_293	63.55
chr4	15693789	15693921	MACS_peak_294	52.85
chr4	22608631	22608780	MACS_peak_295	51.54
chr4	22753260	22753411	MACS_peak_296	51.04
chr4	24324002	24324155	MACS_peak_297	50.56

chr4	37960159	37960296	MACS_peak_298	54.68
chr4	42243243	42243405	MACS_peak_299	54.55
chr4	44899034	44899163	MACS_peak_300	54.76
chr4	107068512	107068647	MACS_peak_301	55.24
chr4	107780048	107780179	MACS_peak_302	126.22
chr4	111515194	111515341	MACS_peak_303	146.58
chr4	124923423	124923565	MACS_peak_304	53.34
chr4	125412096	125412237	MACS_peak_305	62.28
chr4	135784553	135784705	MACS_peak_306	50.8
chr4	141921104	141921287	MACS_peak_307	98
chr4	154165431	154165649	MACS_peak_308	92.41
chr4	156935036	156935185	MACS_peak_309	138.81
chr4	162987409	162987538	MACS_peak_310	138.43
chr4	165475655	165475808	MACS_peak_311	54.98
chr5	5775093	5775241	MACS_peak_312	69.1
chr5	14954335	14954481	MACS_peak_313	76.52
chr5	27679044	27679239	MACS_peak_314	62.39
chr5	32370344	32370481	MACS_peak_315	51.34
chr5	38134421	38134572	MACS_peak_316	70.79
chr5	38760499	38760723	MACS_peak_317	55.87
chr5	46395441	46395548	MACS_peak_318	50.06
chr5	53295804	53295956	MACS_peak_319	50.07
chr5	58964592	58964758	MACS_peak_320	64.92
chr5	62426022	62426162	MACS_peak_321	67.29
chr5	67919497	67919645	MACS_peak_322	50.7
chr5	69300810	69301043	MACS_peak_323	58.3
chr5	69700884	69701026	MACS_peak_324	53.34
chr5	70498941	70499114	MACS_peak_325	55.25
chr5	84229063	84229218	MACS_peak_326	71.32
chr5	88810864	88810989	MACS_peak_327	50.54
chr5	93018816	93018997	MACS_peak_328	69.48
chr5	93644846	93644980	MACS_peak_329	53.69
chr5	101678574	101678696	MACS_peak_330	52.2
chr5	104373249	104373388	MACS_peak_331	53.04
chr5	108402922	108403082	MACS_peak_332	57.78
chr5	109860983	109861122	MACS_peak_333	55.62
chr5	122104157	122104316	MACS_peak_334	82.3
chr5	126582954	126583244	MACS_peak_335	102.07
chr5	131820051	131820169	MACS_peak_336	60.37
chr5	131822639	131822808	MACS_peak_337	139.75
chr5	138003541	138003723	MACS_peak_338	71.71
chr5	144517031	144517172	MACS_peak_339	66.97
chr5	148977825	148977951	MACS_peak_340	56.01

chr5	150655974	150656122	MACS_peak_341	60.26
chr5	158920905	158921050	MACS_peak_342	52.55
chr5	163113850	163113991	MACS_peak_343	50.06
chr5	169022800	169022957	MACS_peak_344	56.22
chr5	174005802	174005966	MACS_peak_345	55.62
chr6	371243	371416	MACS_peak_346	50.56
chr6	3550384	3550601	MACS_peak_347	110.78
chr6	4194166	4194311	MACS_peak_348	52.55
chr6	22339747	22339939	MACS_peak_349	73.84
chr6	26753379	26753554	MACS_peak_350	55.62
chr6	26770384	26770566	MACS_peak_351	71.71
chr6	29690555	29690685	MACS_peak_352	56.67
chr6	34605608	34605684	MACS_peak_353	101.99
chr6	42091125	42091232	MACS_peak_354	105.72
chr6	50675936	50676092	MACS_peak_355	55.62
chr6	58140757	58140966	MACS_peak_356	87.71
chr6	58140971	58141101	MACS_peak_357	61.7
chr6	70188493	70188640	MACS_peak_358	52.04
chr6	76270549	76270685	MACS_peak_359	54.96
chr6	83961901	83962031	MACS_peak_360	84.82
chr6	103456409	103456549	MACS_peak_361	53.87
chr6	103653035	103653179	MACS_peak_362	152.55
chr6	110833742	110833884	MACS_peak_363	61.98
chr6	113594376	113594502	MACS_peak_364	55.62
chr6	120507343	120507467	MACS_peak_365	188.86
chr6	145980207	145980361	MACS_peak_366	50.31
chr6	157731800	157732015	MACS_peak_367	53.84
chr6_cox_hap2	1206726	1206867	MACS_peak_368	71.32
chr6_dbb_hap3	991403	991567	MACS_peak_369	91.57
chr6_qbl_hap6	839460	839585	MACS_peak_370	58.16
chr7	945025	945550	MACS_peak_371	52.25
chr7	6260879	6261036	MACS_peak_372	55.44
chr7	28584860	28585004	MACS_peak_373	52.45
chr7	34321333	34321486	MACS_peak_374	65.8
chr7	37274362	37274501	MACS_peak_375	50.59
chr7	37887780	37888020	MACS_peak_376	194.61
chr7	44214376	44214534	MACS_peak_377	50.18
chr7	44696423	44696578	MACS_peak_378	55.2
chr7	47428601	47428773	MACS_peak_379	148.94
chr7	50160343	50160506	MACS_peak_380	62.09
chr7	51607307	51607457	MACS_peak_381	50.31
chr7	57896476	57896686	MACS_peak_382	55.55
chr7	61365924	61366040	MACS_peak_383	58.9

chr7	61393021	61393100	MACS_peak_384	51.17
chr7	61653916	61654109	MACS_peak_385	61.02
chr7	66560550	66560769	MACS_peak_386	59.19
chr7	71365523	71365675	MACS_peak_387	50.8
chr7	71810257	71810394	MACS_peak_388	54.68
chr7	84320343	84320638	MACS_peak_389	239.48
chr7	85795342	85795479	MACS_peak_390	75.16
chr7	88800593	88800729	MACS_peak_391	71.71
chr7	100937530	100937680	MACS_peak_392	51.29
chr7	115385165	115385294	MACS_peak_393	50.75
chr7	117160187	117160326	MACS_peak_394	51.27
chr7	121928425	121928648	MACS_peak_395	132.72
chr7	132973915	132974063	MACS_peak_396	54.88
chr7	138844736	138844900	MACS_peak_397	50.05
chr7	139062459	139062609	MACS_peak_398	51.29
chr7	144799600	144799825	MACS_peak_399	72.42
chr7	148226473	148226621	MACS_peak_400	51.79
chr7	150003791	150003948	MACS_peak_401	57.82
chr7	154827596	154827769	MACS_peak_402	71.71
chr7	154891069	154891204	MACS_peak_403	55.62
chr8	17321274	17321406	MACS_peak_404	116.78
chr8	17733031	17733186	MACS_peak_405	50.08
chr8	18377752	18377900	MACS_peak_406	51.79
chr8	22215963	22216107	MACS_peak_407	52.81
chr8	23565988	23566091	MACS_peak_408	118.63
chr8	23934016	23934171	MACS_peak_409	50.08
chr8	24440679	24440740	MACS_peak_410	154.17
chr8	43258371	43258606	MACS_peak_411	51.34
chr8	52585374	52585555	MACS_peak_412	52.42
chr8	57162132	57162235	MACS_peak_413	67.43
chr8	68519317	68519491	MACS_peak_414	100.28
chr8	69122345	69122498	MACS_peak_415	50.56
chr8	73123811	73123986	MACS_peak_416	79.95
chr8	79597930	79598094	MACS_peak_417	56.03
chr8	80947326	80947479	MACS_peak_418	50.56
chr8	82158314	82158418	MACS_peak_419	61.93
chr8	99151589	99151750	MACS_peak_420	54.42
chr8	106220103	106220221	MACS_peak_421	54.08
chr8	106416125	106416363	MACS_peak_422	87.71
chr8	108110508	108110665	MACS_peak_423	69.31
chr8	111551860	111552031	MACS_peak_424	53.08
chr8	120398944	120399085	MACS_peak_425	59.09
chr8	125877552	125877685	MACS_peak_426	167.34

chr8	137366019	137366161	MACS_peak_427	136.44
chr8	141913538	141913695	MACS_peak_428	51.79
chr9	10220	10419	MACS_peak_429	113.13
chr9	4773704	4773864	MACS_peak_430	52.61
chr9	8367874	8368018	MACS_peak_431	51.63
chr9	13174279	13174434	MACS_peak_432	50.01
chr9	17675161	17675269	MACS_peak_433	56.79
chr9	20592857	20592983	MACS_peak_434	59.91
chr9	31009818	31009969	MACS_peak_435	51.04
chr9	36817242	36817376	MACS_peak_436	55.52
chr9	43137314	43137484	MACS_peak_437	53.77
chr9	66848819	66849219	MACS_peak_438	50.86
chr9	72458449	72458620	MACS_peak_439	98.59
chr9	78431698	78431844	MACS_peak_440	60.82
chr9	85871682	85871821	MACS_peak_441	51.96
chr9	86641099	86641399	MACS_peak_442	194.29
chr9	96927064	96927196	MACS_peak_443	56.09
chr9	99034023	99034166	MACS_peak_444	53.07
chr9	104284134	104284251	MACS_peak_445	53.77
chr9	109012180	109012425	MACS_peak_446	119.91
chr9	115458009	115458171	MACS_peak_447	178.1
chr9	130297026	130297181	MACS_peak_448	50.08
chr9	131883006	131883161	MACS_peak_449	50.08
chr9	132210386	132210562	MACS_peak_450	51.98
chr9	132541838	132541974	MACS_peak_451	51.4
chr9	139526916	139527060	MACS_peak_452	52.81
chr9_gl000198_random	34841	35009	MACS_peak_453	56.76
chrUn_gl000216	135922	136056	MACS_peak_454	64.27
chrUn_gl000228	96149	96310	MACS_peak_455	55.62
chrUn_gl000239	1151	1320	MACS_peak_456	53.43
chrX	9050363	9050545	MACS_peak_457	113.85
chrX	51364354	51364542	MACS_peak_458	129.59
chrX	61802060	61802155	MACS_peak_459	53.17
chrX	66672096	66672319	MACS_peak_460	66.07
chrX	73618203	73618382	MACS_peak_461	178.02
chrX	81316128	81316262	MACS_peak_462	64.42
chrX	85801325	85801407	MACS_peak_463	96.31
chrX	101667399	101667570	MACS_peak_464	141.51
chrX	104006498	104006652	MACS_peak_465	50.31
chrX	114965499	114965742	MACS_peak_466	66.08
chrX	115000478	115000732	MACS_peak_467	51.99
chrX	115003347	115003553	MACS_peak_468	61
chrX	116917383	116917530	MACS_peak_469	52.04

chrX	117365881	117366035	MACS_peak_470	50.31
chrX	149993010	149993193	MACS_peak_471	98
chrX	153900475	153900705	MACS_peak_472	106.55
chrY	8333485	8333608	MACS_peak_473	108.53
chrY	10075627	10075787	MACS_peak_474	51.34
chrY	13334749	13334874	MACS_peak_475	58.16
chrY	20213011	20213084	MACS_peak_476	80.2
chrY	26170424	26170591	MACS_peak_477	54.29
chrY	28794153	28794254	MACS_peak_478	65.27

Appendix IV: IRF-1 biological replicate 1 peaks generated by ChIP-seq

Chromo	*Start*	*End*	*Peak Identifier*	*Peak Size*
chr1	2584487	2584945	MACS_peak_1	81.35
chr1	2585089	2585963	MACS_peak_2	84
chr1	2586211	2586726	MACS_peak_3	69.28
chr1	2586786	2587510	MACS_peak_4	66.79
chr1	2589205	2589803	MACS_peak_5	61.9
chr1	2600663	2601075	MACS_peak_6	63.06
chr1	2604345	2604604	MACS_peak_7	71.04
chr1	2617049	2617263	MACS_peak_8	63.08
chr1	2617790	2617938	MACS_peak_9	50.35
chr1	2618766	2618984	MACS_peak_10	68.6
chr1	2622716	2622825	MACS_peak_11	72.18
chr1	2623445	2623590	MACS_peak_12	86.03
chr1	2629041	2629412	MACS_peak_13	77.63
chr1	2689032	2689236	MACS_peak_14	109.3
chr1	38009267	38009448	MACS_peak_15	75.28
chr1	121394481	121395235	MACS_peak_16	79.74
chr1	121414920	121415983	MACS_peak_17	168.85
chr1	121422610	121423200	MACS_peak_18	63.13
chr1	121450624	121451455	MACS_peak_19	51.09
chr1	142558313	142558554	MACS_peak_20	50.67
chr1	142782896	142783498	MACS_peak_21	56.51
chr1	143267335	143268131	MACS_peak_22	59.98
chr1	143424889	143425329	MACS_peak_23	96.05
chr1	144196753	144197041	MACS_peak_24	50.28
chr1	145351315	145351973	MACS_peak_25	64.82
chr1	146057062	146057220	MACS_peak_26	61.61
chr1	148595369	148595754	MACS_peak_27	59.82
chr1	148816883	148817179	MACS_peak_28	52
chr1	148823733	148824030	MACS_peak_29	59.14
chr1	148831125	148832144	MACS_peak_30	103.64
chr1	149316584	149316962	MACS_peak_31	59.9
chr1	152740984	152741133	MACS_peak_32	50.26
chr1	157230636	157230818	MACS_peak_33	52.26
chr1	160900978	160901196	MACS_peak_34	76.69
chr1	161422186	161422652	MACS_peak_35	51.53
chr1	172639757	172640002	MACS_peak_36	51.48
chr1	177764438	177764646	MACS_peak_37	50.31
chr1	182779548	182779823	MACS_peak_38	51.67
chr1	191046476	191046731	MACS_peak_39	52.6

chr1	200995903	200996068	MACS_peak_40	61.26
chr1	206509947	206510165	MACS_peak_41	51.55
chr1	228746634	228747537	MACS_peak_42	55.25
chr1	228773488	228774612	MACS_peak_43	74.76
chr1	235810606	235810754	MACS_peak_44	53.29
chr1	245720917	245721070	MACS_peak_45	51.37
chr10	38975086	38975381	MACS_peak_46	52.17
chr10	42643363	42643876	MACS_peak_47	52.14
chr10	90697167	90697331	MACS_peak_48	52
chr10	119759839	119760005	MACS_peak_49	50.88
chr10	131945884	131946051	MACS_peak_50	51.82
chr10	135478733	135478893	MACS_peak_51	54.4
chr10	135499865	135501466	MACS_peak_52	63.4
chr11	642729	642851	MACS_peak_53	78.09
chr11	3674983	3676098	MACS_peak_54	115.06
chr11	35270220	35270386	MACS_peak_55	90.32
chr11	54831615	54831860	MACS_peak_56	51.89
chr11	112282096	112282242	MACS_peak_57	54.08
chr11	126658127	126658280	MACS_peak_58	51.37
chr11	126912028	126912203	MACS_peak_59	50.29
chr12	10517629	10517808	MACS_peak_60	53.04
chr12	38254111	38254335	MACS_peak_61	68.8
chr12	69222543	69222724	MACS_peak_62	108.07
chr12	69229579	69229776	MACS_peak_63	106.26
chr12	69233087	69233623	MACS_peak_64	345.68
chr12	123192144	123192315	MACS_peak_65	52.11
chr13	19089705	19089942	MACS_peak_66	52.96
chr13	67872531	67872718	MACS_peak_67	51.66
chr13	83313127	83313332	MACS_peak_68	64.74
chr13	89112612	89112905	MACS_peak_69	66.77
chr13	112851394	112851507	MACS_peak_70	71.37
chr13	113917103	113917317	MACS_peak_71	89.22
chr14	62334685	62334872	MACS_peak_72	55.54
chr14	64566814	64566949	MACS_peak_73	52.64
chr14	97175295	97175598	MACS_peak_74	124.83
chr15	20025469	20025975	MACS_peak_75	64.63
chr15	20130406	20130887	MACS_peak_76	93.15
chr15	20133937	20134291	MACS_peak_77	53.78
chr15	20546154	20546522	MACS_peak_78	72.44
chr15	21315830	21315972	MACS_peak_79	81.34
chr15	30834151	30834322	MACS_peak_80	55.92
chr15	57634961	57635145	MACS_peak_81	51.33
chr15	75368247	75368399	MACS_peak_82	51.74

chr15	77418014	77418201	MACS_peak_83	63.1
chr16	15322453	15322681	MACS_peak_84	62.9
chr16	32291703	32291862	MACS_peak_85	51.63
chr16	33521732	33522575	MACS_peak_86	51.95
chr16	33934707	33935048	MACS_peak_87	50.84
chr16	34015752	34016560	MACS_peak_88	55.64
chr16	46482653	46483139	MACS_peak_89	96.2
chr16	54126284	54126449	MACS_peak_90	52.54
chr16	83812327	83812534	MACS_peak_91	56.59
chr16	86006195	86006585	MACS_peak_92	56.18
chr16	87479671	87479822	MACS_peak_93	52.13
chr16	90183123	90183991	MACS_peak_94	52.55
chr17	16066646	16066921	MACS_peak_95	52.17
chr17	20147253	20147393	MACS_peak_96	50.62
chr17	39373878	39374074	MACS_peak_97	52.61
chr17_gl000205_random	78952	80094	MACS_peak_98	54.17
chr18	15224442	15224618	MACS_peak_99	50.85
chr18	44546699	44548014	MACS_peak_100	56.81
chr19	24150676	24151028	MACS_peak_101	83.34
chr19	24582793	24583632	MACS_peak_102	93.52
chr19	36762130	36762475	MACS_peak_103	58.14
chr19	36762651	36762951	MACS_peak_104	65.32
chr19	36780205	36780719	MACS_peak_105	55.4
chr19	36784658	36785068	MACS_peak_106	54.45
chr19	37761167	37761522	MACS_peak_107	71.33
chr19	37767361	37767460	MACS_peak_108	57.96
chr19	37779037	37779217	MACS_peak_109	50.25
chr19	37786476	37786794	MACS_peak_110	101.06
chr19	37793975	37794169	MACS_peak_111	84.02
chr19	47871224	47871444	MACS_peak_112	75.86
chr19	48429966	48430525	MACS_peak_113	61.04
chr19	52467799	52468016	MACS_peak_114	96.44
chr19	52468021	52469485	MACS_peak_115	3100
chr19	52471747	52471997	MACS_peak_116	196.26
chr19	52472246	52472421	MACS_peak_117	240.41
chr19	52477613	52477798	MACS_peak_118	73.68
chr19	52505372	52505526	MACS_peak_119	51.91
chr2	6031239	6031412	MACS_peak_120	53.03
chr2	78600299	78600455	MACS_peak_121	53.82
chr2	90464720	90465174	MACS_peak_122	57.06
chr2	91705988	91706137	MACS_peak_123	52.04
chr2	91893800	91894835	MACS_peak_124	56.22
chr2	96617563	96617898	MACS_peak_125	54.45

chr2	222598387	222598542	MACS_peak_126	50.62
chr20	29638216	29638671	MACS_peak_127	63.3
chr20	38368580	38368770	MACS_peak_128	59.96
chr20	43248292	43248447	MACS_peak_129	50.62
chr20	45417423	45417766	MACS_peak_130	53.64
chr20	59904292	59904436	MACS_peak_131	50.28
chr20	62763653	62763903	MACS_peak_132	55.39
chr21	9467640	9468221	MACS_peak_133	54.6
chr21	9825437	9826652	MACS_peak_134	127.19
chr21	9990736	9990989	MACS_peak_135	68.48
chr21	10128641	10129035	MACS_peak_136	53.17
chr21	15254189	15254369	MACS_peak_137	63.89
chr22	17911009	17911211	MACS_peak_138	50.26
chr22	20691398	20691791	MACS_peak_139	68.86
chr3	19602314	19602483	MACS_peak_140	50.9
chr3	75718258	75718584	MACS_peak_141	71.61
chr3	75759501	75759667	MACS_peak_142	75.94
chr3	125676488	125676625	MACS_peak_143	51.82
chr3	191908235	191908421	MACS_peak_144	50.66
chr3	197845033	197845187	MACS_peak_145	50.99
chr4	3590793	3591082	MACS_peak_146	54.02
chr4	9268122	9268345	MACS_peak_147	58.58
chr4	9333329	9333625	MACS_peak_148	57.31
chr4	9368751	9368974	MACS_peak_149	59.84
chr4	49293973	49294367	MACS_peak_150	54.9
chr4	49314749	49316308	MACS_peak_151	72.37
chr4	49323159	49324567	MACS_peak_152	53.44
chr4	49518570	49518674	MACS_peak_153	52.66
chr4	91597509	91597658	MACS_peak_154	56.23
chr4	165025268	165025420	MACS_peak_155	50.88
chr4	186988288	186988455	MACS_peak_156	50.35
chr4	186988522	186988688	MACS_peak_157	54.77
chr4	186990229	186990401	MACS_peak_158	55.94
chr4	190641318	190641879	MACS_peak_159	54.68
chr4	190855836	190856458	MACS_peak_160	59.48
chr4	190995744	190995955	MACS_peak_161	63.54
chr4_gl000193_random	104486	104662	MACS_peak_162	51.08
chr5	17519521	17520808	MACS_peak_163	83.78
chr5	17583665	17585835	MACS_peak_164	58.26
chr5	17589776	17590189	MACS_peak_165	72.12
chr5	34179643	34180743	MACS_peak_166	60.81
chr5	49546619	49546974	MACS_peak_167	52.14
chr5	93018776	93019043	MACS_peak_168	72.15

chr5	131822633	131822841	MACS_peak_169	134.68
chr5	145376954	145377101	MACS_peak_170	53.68
chr5	152146926	152147039	MACS_peak_171	51.39
chr5	157772970	157773135	MACS_peak_172	50.5
chr6	26871403	26871523	MACS_peak_173	55.54
chr6	57607344	57607694	MACS_peak_174	64.63
chr6	157733056	157733212	MACS_peak_175	54.68
chr6	161063193	161063530	MACS_peak_176	50.89
chr6	161065396	161065721	MACS_peak_177	53.1
chr7	20241	20606	MACS_peak_178	111.3
chr7	159812	160007	MACS_peak_179	72.18
chr7	23246323	23246476	MACS_peak_180	51.37
chr7	30818819	30818984	MACS_peak_181	57.86
chr7	32377372	32377518	MACS_peak_182	50.09
chr7	36398626	36398764	MACS_peak_183	50.59
chr7	37102184	37102328	MACS_peak_184	51
chr7	41394053	41394186	MACS_peak_185	52.63
chr7	57918166	57918414	MACS_peak_186	60.84
chr7	58039822	58039972	MACS_peak_187	55.13
chr7	61092281	61092510	MACS_peak_188	52.29
chr7	61243056	61243598	MACS_peak_189	51.57
chr7	61269270	61269717	MACS_peak_190	56.18
chr7	61828162	61828402	MACS_peak_191	53.78
chr7	83377701	83377969	MACS_peak_192	51.08
chr7	89325916	89326039	MACS_peak_193	51.65
chr7	104201858	104202011	MACS_peak_194	57.04
chr7	120755018	120755188	MACS_peak_195	50.66
chr7	126958307	126958463	MACS_peak_196	50.25
chr7	153343674	153343830	MACS_peak_197	51.13
chr7	157951203	157951414	MACS_peak_198	80.71
chr8	7623326	7623458	MACS_peak_199	53.89
chr8	58118271	58118700	MACS_peak_200	57.59
chr8	58125482	58125879	MACS_peak_201	54.09
chr8	64011996	64012151	MACS_peak_202	50.62
chr8	86559493	86566561	MACS_peak_203	87.06
chr8	86569166	86570636	MACS_peak_204	81.42
chr8	86575781	86576421	MACS_peak_205	76.58
chr8	86738756	86739845	MACS_peak_206	61.59
chr8	86759130	86760345	MACS_peak_207	86.38
chr8	86769874	86771468	MACS_peak_208	63.81
chr8	86789697	86790348	MACS_peak_209	55.04
chr8	86806578	86807529	MACS_peak_210	56.89
chr8	86809754	86811364	MACS_peak_211	51.62

chr8	86812353	86813295	MACS_peak_212	68.48
chr8	86836743	86837909	MACS_peak_213	78.52
chr8	87296288	87296434	MACS_peak_214	50.77
chr8	103313982	103314139	MACS_peak_215	56.3
chr8	129389388	129389591	MACS_peak_216	78.09
chr8	131870645	131870831	MACS_peak_217	55.53
chr8	142502371	142502517	MACS_peak_218	60.06
chr9	8014257	8014433	MACS_peak_219	59.12
chr9	15291433	15291768	MACS_peak_220	69.11
chr9	42079495	42079734	MACS_peak_221	55.39
chr9	42523995	42524258	MACS_peak_222	51.69
chr9	44123829	44124001	MACS_peak_223	69.83
chr9	44393015	44393210	MACS_peak_224	50.46
chr9	68346640	68346916	MACS_peak_225	55.38
chr9	69051254	69051993	MACS_peak_226	56.02
chr9	93970433	93970596	MACS_peak_227	52.37
chr9	104072700	104072851	MACS_peak_228	52.71
chr9	110008675	110008840	MACS_peak_229	58.21
chr9	122707265	122707404	MACS_peak_230	51.02
chr9	139526878	139527052	MACS_peak_231	78.55
chrUn_gl000211	54714	54911	MACS_peak_232	55.58
chrUn_gl000217	127487	127850	MACS_peak_233	94.5
chrUn_gl000219	148053	148353	MACS_peak_234	69.85
chrUn_gl000224	46950	47680	MACS_peak_235	50.85
chrUn_gl000228	18907	19152	MACS_peak_236	57.35
chrUn_gl000228	106086	106286	MACS_peak_237	85.46
chrUn_gl000232	25001	25438	MACS_peak_238	56.37
chrUn_gl000235	9514	9785	MACS_peak_239	52.21
chrUn_gl000243	31884	32322	MACS_peak_240	51.7
chrX	9378960	9379250	MACS_peak_241	50.84
chrX	9380649	9380846	MACS_peak_242	71.57
chrX	49355548	49355702	MACS_peak_243	53.68
chrX	61833550	61833838	MACS_peak_244	65.09
chrX	61866432	61866915	MACS_peak_245	64.4
chrX	61896803	61897093	MACS_peak_246	65.04
chrX	134860270	134860429	MACS_peak_247	53.96
chrX	134962851	134962999	MACS_peak_248	54.68
chrY	9921770	9921997	MACS_peak_249	60.8
chrY	9937881	9938149	MACS_peak_250	57.78
chrY	9981708	9981863	MACS_peak_251	50.62
chrY	10017856	10017995	MACS_peak_252	51.02
chrY	10065026	10065191	MACS_peak_253	52.53
chrY	10102993	10103068	MACS_peak_254	143.75

chrY	13419859	13420247	MACS_peak_255	54.2
chrY	13496799	13496956	MACS_peak_256	61.44
chrY	28798841	28798923	MACS_peak_257	85.01
chrY	28809285	28809366	MACS_peak_258	57.4

Appendix V: IRF-1 biological replicate 2 peaks generated by ChIP-seq

Chromo	*Start*	*End*	*Peak Identifier*	*Peak Size*
chr1	6356710	6356856	MACS_peak_1	50.72
chr1	12243933	12244057	MACS_peak_2	56.95
chr1	14121644	14121791	MACS_peak_3	51.6
chr1	15875115	15875262	MACS_peak_4	61.43
chr1	17081539	17081817	MACS_peak_5	51.41
chr1	17630281	17630426	MACS_peak_6	52.03
chr1	19326383	19326527	MACS_peak_7	52.24
chr1	24689980	24690120	MACS_peak_8	50.16
chr1	28242442	28242566	MACS_peak_9	56.95
chr1	30961553	30961675	MACS_peak_10	57.46
chr1	35781108	35781252	MACS_peak_11	50.45
chr1	54010740	54010888	MACS_peak_12	51.39
chr1	62484331	62484504	MACS_peak_13	55.71
chr1	62915792	62915933	MACS_peak_14	51.39
chr1	63351013	63351166	MACS_peak_15	52.25
chr1	74935602	74935758	MACS_peak_16	57.42
chr1	81729188	81729360	MACS_peak_17	50.72
chr1	97174381	97174510	MACS_peak_18	62.69
chr1	104107943	104108082	MACS_peak_19	52.15
chr1	104760995	104761167	MACS_peak_20	58.47
chr1	106227465	106227606	MACS_peak_21	51.17
chr1	107470221	107470369	MACS_peak_22	52.05
chr1	108553848	108553991	MACS_peak_23	52.46
chr1	116154952	116155092	MACS_peak_24	63.16
chr1	118098703	118098857	MACS_peak_25	50.72
chr1	120755615	120755847	MACS_peak_26	56.9
chr1	121353841	121354636	MACS_peak_27	56.34
chr1	121354974	121355371	MACS_peak_28	81.56
chr1	121355697	121355805	MACS_peak_29	51.22
chr1	121356894	121357446	MACS_peak_30	50.63
chr1	121359835	121360259	MACS_peak_31	53.11
chr1	121363270	121363541	MACS_peak_32	58.74
chr1	121394549	121395221	MACS_peak_33	71.86
chr1	121406212	121406473	MACS_peak_34	76.59
chr1	121450693	121450864	MACS_peak_35	60.93
chr1	121462457	121462627	MACS_peak_36	61.11
chr1	121468069	121468220	MACS_peak_37	51.27
chr1	121484049	121485439	MACS_peak_38	445.89
chr1	142956783	142957113	MACS_peak_39	54.39

chr1	143123124	143123317	MACS_peak_40	52.21
chr1	143972447	143972688	MACS_peak_41	67.14
chr1	144209036	144209196	MACS_peak_42	50.14
chr1	144616539	144616769	MACS_peak_43	51.28
chr1	147578485	147578590	MACS_peak_44	50.09
chr1	148275738	148276045	MACS_peak_45	60.34
chr1	148321904	148322047	MACS_peak_46	50.58
chr1	148910242	148910549	MACS_peak_47	52.9
chr1	149016058	149016306	MACS_peak_48	77.32
chr1	149076674	149076845	MACS_peak_49	52.21
chr1	150274464	150274608	MACS_peak_50	70.23
chr1	151787790	151788061	MACS_peak_51	101.64
chr1	156588876	156589032	MACS_peak_52	50.72
chr1	157759557	157759727	MACS_peak_53	51.41
chr1	161200085	161200223	MACS_peak_54	50.33
chr1	165606142	165606267	MACS_peak_55	50.21
chr1	168577602	168577723	MACS_peak_56	50.47
chr1	169984098	169984218	MACS_peak_57	56.25
chr1	174348718	174348850	MACS_peak_58	51.43
chr1	174795977	174796090	MACS_peak_59	56.88
chr1	182381983	182382159	MACS_peak_60	59.68
chr1	184550773	184550908	MACS_peak_61	56.74
chr1	185314968	185315101	MACS_peak_62	51.19
chr1	195519234	195519378	MACS_peak_63	54.07
chr1	199494837	199494981	MACS_peak_64	50.7
chr1	208811801	208812008	MACS_peak_65	63.04
chr1	214459198	214459351	MACS_peak_66	63.64
chr1	220660919	220661088	MACS_peak_67	51.23
chr1	224729436	224729568	MACS_peak_68	50.72
chr1	232896510	232896627	MACS_peak_69	53.3
chr1	236043557	236043672	MACS_peak_70	56.32
chr1	242881323	242881476	MACS_peak_71	62.98
chr1	244722973	244723167	MACS_peak_72	60.2
chr1	246174690	246174818	MACS_peak_73	57.08
chr1	246407911	246408003	MACS_peak_74	61.93
chr1	247213690	247213825	MACS_peak_75	59.68
chr10	2126560	2126704	MACS_peak_76	62.16
chr10	14868013	14868105	MACS_peak_77	60.52
chr10	23424743	23424897	MACS_peak_78	59.79
chr10	25722113	25722258	MACS_peak_79	52.03
chr10	26042300	26042436	MACS_peak_80	60.82
chr10	35306101	35306251	MACS_peak_81	50.36
chr10	38992331	38992448	MACS_peak_82	58.79

chr10	39047848	39047971	MACS_peak_83	70.9
chr10	42359798	42361279	MACS_peak_84	56.61
chr10	42383236	42384159	MACS_peak_85	119.44
chr10	42384902	42386257	MACS_peak_86	149.07
chr10	42386925	42389785	MACS_peak_87	578.29
chr10	42390095	42395350	MACS_peak_88	610.11
chr10	42395362	42397376	MACS_peak_89	250.77
chr10	42398310	42399106	MACS_peak_90	138.67
chr10	42400092	42400809	MACS_peak_91	203.18
chr10	42409113	42409475	MACS_peak_92	71.91
chr10	42527892	42528175	MACS_peak_93	73.26
chr10	42529202	42530391	MACS_peak_94	227.56
chr10	42533779	42534238	MACS_peak_95	85.27
chr10	42534363	42535068	MACS_peak_96	100.96
chr10	42535092	42535285	MACS_peak_97	50.05
chr10	42535720	42536135	MACS_peak_98	51.39
chr10	42536548	42537077	MACS_peak_99	74.81
chr10	42538523	42539231	MACS_peak_100	70.1
chr10	42539883	42540354	MACS_peak_101	124.39
chr10	42541298	42542209	MACS_peak_102	122.1
chr10	42542692	42543452	MACS_peak_103	99.97
chr10	42543537	42543952	MACS_peak_104	53
chr10	42544356	42545134	MACS_peak_105	105.83
chr10	42545779	42546472	MACS_peak_106	106.85
chr10	42596687	42599050	MACS_peak_107	363.64
chr10	42599065	42599438	MACS_peak_108	125.59
chr10	42599488	42600586	MACS_peak_109	285.63
chr10	49244166	49244288	MACS_peak_110	57.46
chr10	49995928	49996064	MACS_peak_111	57.76
chr10	53273133	53273246	MACS_peak_112	56.88
chr10	81454074	81454266	MACS_peak_113	51.44
chr10	83445681	83445815	MACS_peak_114	51.52
chr10	84032615	84032754	MACS_peak_115	53.35
chr10	87424152	87424293	MACS_peak_116	52.9
chr10	88163251	88163470	MACS_peak_117	104.51
chr10	90897032	90897171	MACS_peak_118	50.38
chr10	94799930	94800067	MACS_peak_119	60.56
chr10	95361761	95361918	MACS_peak_120	59.11
chr10	98599810	98599961	MACS_peak_121	57.14
chr10	100948305	100948453	MACS_peak_122	50.72
chr10	101093726	101093874	MACS_peak_123	51.39
chr10	101900001	101900182	MACS_peak_124	83.48
chr10	107326183	107326315	MACS_peak_125	54.97

chr10	109607387	109607487	MACS_peak_126	58.44
chr10	112348014	112348144	MACS_peak_127	51.35
chr10	112719945	112720075	MACS_peak_128	55.45
chr10	113133687	113133832	MACS_peak_129	52.03
chr10	117255494	117255634	MACS_peak_130	50.16
chr10	118538245	118538370	MACS_peak_131	50.98
chr10	122050011	122050144	MACS_peak_132	54.73
chr10	129382921	129383085	MACS_peak_133	75.08
chr10	130115209	130115350	MACS_peak_134	52.9
chr11	25078246	25078360	MACS_peak_135	59.62
chr11	29490839	29490956	MACS_peak_136	55.78
chr11	33292930	33293137	MACS_peak_137	55.79
chr11	33819546	33819651	MACS_peak_138	58.64
chr11	35401154	35401279	MACS_peak_139	53.69
chr11	35702604	35702739	MACS_peak_140	53.04
chr11	39825526	39825673	MACS_peak_141	50.24
chr11	42862333	42862510	MACS_peak_142	60.71
chr11	48932315	48932547	MACS_peak_143	52.71
chr11	50687892	50688104	MACS_peak_144	56.71
chr11	50695169	50695293	MACS_peak_145	50.96
chr11	50739010	50739097	MACS_peak_146	77.11
chr11	51319350	51319499	MACS_peak_147	51.18
chr11	51580759	51581126	MACS_peak_148	70.18
chr11	51582416	51583595	MACS_peak_149	73.4
chr11	51586670	51587499	MACS_peak_150	174.89
chr11	51588237	51588689	MACS_peak_151	50.21
chr11	51588702	51589538	MACS_peak_152	192.68
chr11	51590282	51591422	MACS_peak_153	225.25
chr11	51592640	51593455	MACS_peak_154	64.62
chr11	55820040	55820158	MACS_peak_155	56.09
chr11	56484991	56485121	MACS_peak_156	52.75
chr11	58199036	58199163	MACS_peak_157	61.98
chr11	58813429	58813562	MACS_peak_158	52.91
chr11	64041981	64042126	MACS_peak_159	51.42
chr11	68316117	68316239	MACS_peak_160	54.46
chr11	73279063	73279240	MACS_peak_161	52.09
chr11	73671255	73671384	MACS_peak_162	55.7
chr11	78393658	78393801	MACS_peak_163	52.46
chr11	80726839	80726984	MACS_peak_164	50.81
chr11	84707457	84707607	MACS_peak_165	50.97
chr11	85151333	85151466	MACS_peak_166	51.76
chr11	89124096	89124223	MACS_peak_167	50.49
chr11	91400918	91401066	MACS_peak_168	50.78

chr11	98303310	98303476	MACS_peak_169	52.02
chr11	99827336	99827503	MACS_peak_170	53.64
chr11	100109649	100109800	MACS_peak_171	50.77
chr11	103258645	103258771	MACS_peak_172	52.05
chr11	106819721	106819849	MACS_peak_173	50.24
chr11	111921024	111921177	MACS_peak_174	60.02
chr11	112580049	112580253	MACS_peak_175	91.49
chr11	117104302	117104457	MACS_peak_176	50.72
chr11	118311595	118311745	MACS_peak_177	50.97
chr11	123516868	123517020	MACS_peak_178	50.72
chr11	132817009	132817130	MACS_peak_179	57.72
chr11	134097147	134097274	MACS_peak_180	56.19
chr12	2629355	2629578	MACS_peak_181	92.73
chr12	2648635	2649043	MACS_peak_182	89.44
chr12	3503972	3504104	MACS_peak_183	53.15
chr12	7706784	7706908	MACS_peak_184	55.71
chr12	9848506	9848624	MACS_peak_185	59.68
chr12	16419771	16419907	MACS_peak_186	52.81
chr12	22663233	22663385	MACS_peak_187	65.24
chr12	29673016	29673115	MACS_peak_188	56.58
chr12	34347363	34347490	MACS_peak_189	60.2
chr12	34824812	34824945	MACS_peak_190	50.92
chr12	38418730	38418862	MACS_peak_191	54.97
chr12	39230738	39230863	MACS_peak_192	56.69
chr12	44059851	44059995	MACS_peak_193	51.63
chr12	48990743	48990887	MACS_peak_194	52.24
chr12	53821341	53821492	MACS_peak_195	60.48
chr12	55272732	55272899	MACS_peak_196	50.72
chr12	63515333	63515490	MACS_peak_197	55.78
chr12	67722767	67722901	MACS_peak_198	52.09
chr12	68336888	68337035	MACS_peak_199	51.6
chr12	69202978	69203116	MACS_peak_200	52.21
chr12	69210588	69210733	MACS_peak_201	94.6
chr12	69222551	69222707	MACS_peak_202	148.53
chr12	69229613	69229761	MACS_peak_203	203.99
chr12	69230456	69230555	MACS_peak_204	60.02
chr12	69233055	69233655	MACS_peak_205	489.4
chr12	75956289	75956470	MACS_peak_206	52.94
chr12	76371138	76371253	MACS_peak_207	55.75
chr12	84417368	84417506	MACS_peak_208	50.72
chr12	95825707	95825820	MACS_peak_209	53.36
chr12	101904217	101904315	MACS_peak_210	60.14
chr12	108979541	108979660	MACS_peak_211	58.25

chr12	110485717	110485846	MACS_peak_212	55.7
chr12	113150819	113150922	MACS_peak_213	57.06
chr12	115100078	115100193	MACS_peak_214	59.34
chr12	117249427	117249551	MACS_peak_215	67.51
chr12	129201014	129201150	MACS_peak_216	54.03
chr12	129746707	129746822	MACS_peak_217	66.8
chr12	132073718	132073870	MACS_peak_218	51.14
chr13	24552539	24552638	MACS_peak_219	61.32
chr13	25644606	25644755	MACS_peak_220	51.18
chr13	28840329	28840475	MACS_peak_221	56.16
chr13	31147021	31147162	MACS_peak_222	55.07
chr13	34656895	34657027	MACS_peak_223	50.34
chr13	48590817	48591041	MACS_peak_224	71.67
chr13	53027615	53027775	MACS_peak_225	55.12
chr13	56965807	56965919	MACS_peak_226	57.74
chr13	65586964	65587107	MACS_peak_227	56.59
chr13	67916883	67917033	MACS_peak_228	60.72
chr13	71754921	71755076	MACS_peak_229	52.36
chr13	79278885	79279007	MACS_peak_230	50.47
chr13	82792292	82792436	MACS_peak_231	50.72
chr13	90280421	90280540	MACS_peak_232	55.1
chr13	94587204	94587385	MACS_peak_233	50.74
chr13	96056970	96057105	MACS_peak_234	54.26
chr13	96556805	96556955	MACS_peak_235	50.3
chr13	97779851	97780033	MACS_peak_236	61.8
chr13	101845674	101845786	MACS_peak_237	51.67
chr13	102813930	102814060	MACS_peak_238	50.28
chr13	110338771	110338885	MACS_peak_239	58.54
chr13	112934098	112934225	MACS_peak_240	53.48
chr14	19017876	19018039	MACS_peak_241	57.87
chr14	19256264	19256405	MACS_peak_242	50.72
chr14	19606345	19606490	MACS_peak_243	59.68
chr14	19713254	19713404	MACS_peak_244	74.66
chr14	19861972	19862174	MACS_peak_245	64.93
chr14	25000972	25001085	MACS_peak_246	59.9
chr14	40760390	40760533	MACS_peak_247	62.41
chr14	54270515	54270664	MACS_peak_248	51.18
chr14	64012791	64012844	MACS_peak_249	132.11
chr14	71074139	71074206	MACS_peak_250	55.67
chr14	71686234	71686379	MACS_peak_251	52.03
chr14	82329449	82329597	MACS_peak_252	54.81
chr14	88548794	88548944	MACS_peak_253	50.97
chr15	20124978	20125052	MACS_peak_254	62.24

chr15	20510895	20510988	MACS_peak_255	50.36
chr15	22042809	22042957	MACS_peak_256	54.81
chr15	23574447	23574576	MACS_peak_257	53.57
chr15	32518925	32519074	MACS_peak_258	51.18
chr15	32560921	32561062	MACS_peak_259	50.51
chr15	32597339	32597465	MACS_peak_260	52.88
chr15	34471606	34471776	MACS_peak_261	59.68
chr15	34531807	34531948	MACS_peak_262	61.87
chr15	44070438	44070585	MACS_peak_263	51.6
chr15	44644335	44644498	MACS_peak_264	50.72
chr15	46865940	46866088	MACS_peak_265	51.39
chr15	47635883	47635988	MACS_peak_266	62.26
chr15	49552408	49552523	MACS_peak_267	51.56
chr15	55951678	55951858	MACS_peak_268	50.42
chr15	56431404	56431624	MACS_peak_269	58.79
chr15	66123017	66123155	MACS_peak_270	50.89
chr15	69808547	69808694	MACS_peak_271	50.72
chr15	71468126	71468225	MACS_peak_272	63.19
chr15	71678127	71678290	MACS_peak_273	50.72
chr15	73754489	73754630	MACS_peak_274	52.9
chr15	76188435	76188571	MACS_peak_275	54.03
chr15	80666960	80667135	MACS_peak_276	50.72
chr15	83389306	83389430	MACS_peak_277	191.17
chr15	84264976	84265115	MACS_peak_278	50.38
chr15	85852116	85852248	MACS_peak_279	57.37
chr15	93486643	93486781	MACS_peak_280	50.72
chr15	99949125	99949273	MACS_peak_281	51.39
chr15	100098362	100098550	MACS_peak_282	50.21
chr16	3028669	3028794	MACS_peak_283	56.69
chr16	5304657	5304826	MACS_peak_284	53.23
chr16	6748865	6749013	MACS_peak_285	51.39
chr16	7474686	7474882	MACS_peak_286	59.68
chr16	8881002	8881189	MACS_peak_287	59.68
chr16	9825218	9825354	MACS_peak_288	54.03
chr16	11105363	11105495	MACS_peak_289	51.99
chr16	12941843	12941986	MACS_peak_290	50.07
chr16	18196300	18196440	MACS_peak_291	51.61
chr16	18456330	18456479	MACS_peak_292	51.18
chr16	18539034	18539157	MACS_peak_293	56.89
chr16	19146451	19146583	MACS_peak_294	54.97
chr16	21055331	21055463	MACS_peak_295	54.97
chr16	22450735	22450875	MACS_peak_296	51.9
chr16	22970971	22971110	MACS_peak_297	54.21

chr16	26144381	26144525	MACS_peak_298	52.24
chr16	30297587	30297741	MACS_peak_299	54.2
chr16	30930217	30930306	MACS_peak_300	70.36
chr16	31193274	31193415	MACS_peak_301	52.9
chr16	31638396	31638541	MACS_peak_302	60.2
chr16	32789841	32789990	MACS_peak_303	52.21
chr16	33037985	33038118	MACS_peak_304	58.03
chr16	34022083	34022344	MACS_peak_305	75.68
chr16	35239610	35239770	MACS_peak_306	53.74
chr16	46385801	46396897	MACS_peak_307	597.96
chr16	46396926	46402391	MACS_peak_308	657.44
chr16	46402399	46409563	MACS_peak_309	1601.41
chr16	46411680	46414537	MACS_peak_310	331.93
chr16	46418945	46423504	MACS_peak_311	70.29
chr16	46423533	46426550	MACS_peak_312	359.54
chr16	46431965	46435675	MACS_peak_313	111.59
chr16	46440624	46440756	MACS_peak_314	50.07
chr16	48052046	48052196	MACS_peak_315	57.37
chr16	51790368	51790502	MACS_peak_316	51.24
chr16	55771816	55771954	MACS_peak_317	53.57
chr16	55850597	55850719	MACS_peak_318	56.23
chr16	56170470	56170646	MACS_peak_319	63.35
chr16	61849484	61849642	MACS_peak_320	52.55
chr16	64007532	64007727	MACS_peak_321	52.79
chr16	69565802	69565946	MACS_peak_322	50.72
chr16	69719720	69719835	MACS_peak_323	56.32
chr16	71879378	71879521	MACS_peak_324	52.46
chr16	72999428	72999559	MACS_peak_325	61.82
chr16	73036192	73036340	MACS_peak_326	59.44
chr16	74581953	74582122	MACS_peak_327	54.39
chr16	76003085	76003175	MACS_peak_328	79.05
chr16	77709175	77709320	MACS_peak_329	52.03
chr16	78352686	78352854	MACS_peak_330	66.44
chr16	89398958	89399097	MACS_peak_331	50.38
chr17	232568	232719	MACS_peak_332	50.77
chr17	3520835	3520971	MACS_peak_333	54.03
chr17	6829921	6830090	MACS_peak_334	50.72
chr17	7572793	7572943	MACS_peak_335	70.91
chr17	7578377	7578528	MACS_peak_336	50.77
chr17	7579384	7579530	MACS_peak_337	51.81
chr17	17560600	17560741	MACS_peak_338	52.9
chr17	19093517	19093663	MACS_peak_339	51.81
chr17	22148703	22148844	MACS_peak_340	59.54

chr17	22245292	22248927	MACS_peak_341	315.53
chr17	22248950	22249863	MACS_peak_342	104.11
chr17	22249921	22251134	MACS_peak_343	212.49
chr17	22254509	22256982	MACS_peak_344	217.96
chr17	22256986	22258891	MACS_peak_345	343.92
chr17	22258897	22261705	MACS_peak_346	438.32
chr17	22262230	22263005	MACS_peak_347	71.06
chr17	26106405	26106536	MACS_peak_348	55.21
chr17	28660502	28660698	MACS_peak_349	68.65
chr17	29246713	29246839	MACS_peak_350	50.73
chr17	30993001	30993123	MACS_peak_351	57.46
chr17	34197865	34197997	MACS_peak_352	54.97
chr17	34569111	34569261	MACS_peak_353	59.68
chr17	41086434	41086546	MACS_peak_354	59.56
chr17	47860950	47861059	MACS_peak_355	52.75
chr17	55745517	55745646	MACS_peak_356	55.7
chr17	60165669	60165791	MACS_peak_357	311.97
chr17	66762406	66762548	MACS_peak_358	52.95
chr17	71981219	71981364	MACS_peak_359	52.03
chr17	74707986	74708134	MACS_peak_360	51.39
chr18	107812	109429	MACS_peak_361	139
chr18	18512688	18513184	MACS_peak_362	70.31
chr18	18517017	18517784	MACS_peak_363	74.15
chr18	18518712	18520345	MACS_peak_364	183.76
chr18	33075654	33075805	MACS_peak_365	50.77
chr18	41857057	41857196	MACS_peak_366	53.35
chr18	44575411	44575474	MACS_peak_367	74.31
chr19	4727243	4727391	MACS_peak_368	51.39
chr19	5253486	5253596	MACS_peak_369	50.19
chr19	7156339	7156490	MACS_peak_370	50.77
chr19	7470835	7470973	MACS_peak_371	53.57
chr19	11350701	11350852	MACS_peak_372	50.77
chr19	12899892	12900041	MACS_peak_373	51.18
chr19	13942569	13942714	MACS_peak_374	52.03
chr19	16899916	16900040	MACS_peak_375	53.95
chr19	24625445	24626215	MACS_peak_376	68.26
chr19	27731787	27732875	MACS_peak_377	243.63
chr19	27732969	27733654	MACS_peak_378	139.06
chr19	27733691	27734834	MACS_peak_379	208.38
chr19	27736131	27736979	MACS_peak_380	143.85
chr19	27737351	27737840	MACS_peak_381	61.55
chr19	27738376	27738766	MACS_peak_382	140.74
chr19	27739141	27739531	MACS_peak_383	72.03

chr19	27740004	27740302	MACS_peak_384	57.71
chr19	27764106	27764226	MACS_peak_385	69.76
chr19	27805875	27806050	MACS_peak_386	58.79
chr19	36765035	36765231	MACS_peak_387	56.64
chr19	43758640	43758787	MACS_peak_388	50.72
chr19	45128897	45129051	MACS_peak_389	59.79
chr19	47756169	47756317	MACS_peak_390	70.66
chr19	48407660	48407818	MACS_peak_391	52.18
chr19	51684117	51684266	MACS_peak_392	51.18
chr19	52468093	52469467	MACS_peak_393	2197.07
chr19	52471834	52471950	MACS_peak_394	69.92
chr19	53390103	53390203	MACS_peak_395	63.19
chr19	53535026	53535174	MACS_peak_396	51.39
chr1_gl000192_random	203252	203445	MACS_peak_397	53.79
chr1_gl000192_random	355502	355625	MACS_peak_398	55.88
chr2	4779218	4779351	MACS_peak_399	51.76
chr2	5766225	5766390	MACS_peak_400	52.61
chr2	11389455	11389540	MACS_peak_401	68.7
chr2	13477396	13477588	MACS_peak_402	66.73
chr2	19510181	19510319	MACS_peak_403	50.72
chr2	21361631	21361785	MACS_peak_404	59.68
chr2	21899010	21899124	MACS_peak_405	56.6
chr2	25672830	25672981	MACS_peak_406	50.77
chr2	26104746	26104860	MACS_peak_407	51.35
chr2	31907021	31907165	MACS_peak_408	52.99
chr2	39940918	39941044	MACS_peak_409	52.88
chr2	48461013	48461149	MACS_peak_410	50.23
chr2	49445313	49445464	MACS_peak_411	50.77
chr2	59376771	59376908	MACS_peak_412	53.8
chr2	63066272	63066393	MACS_peak_413	53.59
chr2	77821112	77821241	MACS_peak_414	55.7
chr2	79320032	79320180	MACS_peak_415	51.39
chr2	79470512	79470650	MACS_peak_416	53.57
chr2	80121115	80121247	MACS_peak_417	59.68
chr2	80889949	80890094	MACS_peak_418	52.03
chr2	84781968	84782117	MACS_peak_419	50.57
chr2	85222720	85222849	MACS_peak_420	50.72
chr2	89629911	89630070	MACS_peak_421	101.87
chr2	89872004	89872438	MACS_peak_422	64.62
chr2	89875024	89875471	MACS_peak_423	57.82
chr2	89891171	89891263	MACS_peak_424	51.85
chr2	90298434	90298623	MACS_peak_425	57.49
chr2	90415854	90416094	MACS_peak_426	57.11

chr2	92268931	92269906	MACS_peak_427	84.34
chr2	92281243	92281405	MACS_peak_428	65
chr2	92289912	92290328	MACS_peak_429	92.86
chr2	92291324	92291831	MACS_peak_430	114.45
chr2	92295776	92297012	MACS_peak_431	104.2
chr2	92297288	92297594	MACS_peak_432	87.85
chr2	92297648	92298304	MACS_peak_433	100.1
chr2	92300525	92301312	MACS_peak_434	54.67
chr2	92305600	92306667	MACS_peak_435	70.29
chr2	92306712	92308404	MACS_peak_436	377.45
chr2	92308430	92309407	MACS_peak_437	162.39
chr2	92309460	92311359	MACS_peak_438	102.31
chr2	92311829	92314403	MACS_peak_439	482.26
chr2	92314600	92322263	MACS_peak_440	1511.34
chr2	92322293	92322917	MACS_peak_441	145.17
chr2	92322958	92326171	MACS_peak_442	178.63
chr2	97819669	97819820	MACS_peak_443	50.27
chr2	98555333	98555480	MACS_peak_444	51.6
chr2	101383419	101383566	MACS_peak_445	50.72
chr2	103178162	103178311	MACS_peak_446	51.18
chr2	132904711	132904871	MACS_peak_447	54.81
chr2	133012724	133012795	MACS_peak_448	51.44
chr2	133112687	133112889	MACS_peak_449	54.83
chr2	185155895	185156045	MACS_peak_450	50.97
chr2	196241553	196241700	MACS_peak_451	61.43
chr2	196757538	196757682	MACS_peak_452	52.24
chr2	205479101	205479242	MACS_peak_453	52.9
chr2	207694760	207694911	MACS_peak_454	50.77
chr2	221410902	221411207	MACS_peak_455	166.02
chr2	224155420	224155543	MACS_peak_456	51.48
chr2	243160169	243160436	MACS_peak_457	67.08
chr20	2662863	2663002	MACS_peak_458	50.38
chr20	8884611	8884761	MACS_peak_459	50.97
chr20	22136274	22136383	MACS_peak_460	59.5
chr20	22519572	22519668	MACS_peak_461	74.59
chr20	22969450	22969598	MACS_peak_462	51.39
chr20	23269526	23269641	MACS_peak_463	59.34
chr20	23789369	23789482	MACS_peak_464	50.72
chr20	26260256	26260446	MACS_peak_465	50.96
chr20	34218862	34219075	MACS_peak_466	55.06
chr20	36902860	36903062	MACS_peak_467	60.2
chr20	42707348	42707461	MACS_peak_468	56.02
chr20	45069198	45069354	MACS_peak_469	50.72

chr20	45937374	45937545	MACS_peak_470	67.95
chr20	47147285	47147419	MACS_peak_471	51.52
chr20	47281061	47281200	MACS_peak_472	50.38
chr20	47683243	47683362	MACS_peak_473	52.41
chr20	49623083	49623195	MACS_peak_474	71.19
chr20	53036121	53036284	MACS_peak_475	52.58
chr21	9998452	9998562	MACS_peak_476	53.43
chr21	14358972	14359241	MACS_peak_477	67.48
chr21	33932127	33932195	MACS_peak_478	98.71
chr21	41040942	41041091	MACS_peak_479	60.95
chr22	18718027	18718206	MACS_peak_480	72.34
chr22	18883763	18884107	MACS_peak_481	54.05
chr22	19247249	19247380	MACS_peak_482	55.21
chr22	39281983	39282139	MACS_peak_483	59.33
chr3	1824783	1824935	MACS_peak_484	60.25
chr3	1886270	1886420	MACS_peak_485	50.97
chr3	2329329	2329470	MACS_peak_486	61.4
chr3	10557491	10557629	MACS_peak_487	53.57
chr3	13898337	13898481	MACS_peak_488	52.24
chr3	17685518	17685654	MACS_peak_489	51.06
chr3	26357339	26357473	MACS_peak_490	54.5
chr3	29540169	29540301	MACS_peak_491	53.15
chr3	33007757	33007929	MACS_peak_492	55.92
chr3	33363797	33363944	MACS_peak_493	74.22
chr3	34172888	34173003	MACS_peak_494	66.8
chr3	35366686	35366806	MACS_peak_495	52.77
chr3	39563102	39563232	MACS_peak_496	54.22
chr3	51363944	51364067	MACS_peak_497	51.22
chr3	55036131	55036261	MACS_peak_498	55.45
chr3	58294905	58295058	MACS_peak_499	59.89
chr3	58655419	58655475	MACS_peak_500	238.15
chr3	66364673	66364816	MACS_peak_501	56
chr3	66807050	66807206	MACS_peak_502	50.72
chr3	69171842	69171989	MACS_peak_503	51.6
chr3	69538887	69538996	MACS_peak_504	56.89
chr3	71493566	71493712	MACS_peak_505	51.81
chr3	76320586	76320725	MACS_peak_506	52.43
chr3	76871143	76871287	MACS_peak_507	52.24
chr3	78421801	78421948	MACS_peak_508	55.04
chr3	84932205	84932354	MACS_peak_509	51.18
chr3	86195185	86195325	MACS_peak_510	50.16
chr3	86204377	86204511	MACS_peak_511	54.5
chr3	100672403	100672567	MACS_peak_512	50.71

chr3	100908437	100908587	MACS_peak_513	51.59
chr3	112917970	112918104	MACS_peak_514	52.09
chr3	113802126	113802193	MACS_peak_515	86.29
chr3	114332769	114332938	MACS_peak_516	59.68
chr3	117087689	117087813	MACS_peak_517	53.38
chr3	120552019	120552146	MACS_peak_518	53.2
chr3	120696079	120696226	MACS_peak_519	50.99
chr3	126640808	126640946	MACS_peak_520	53.57
chr3	128465535	128465683	MACS_peak_521	51.39
chr3	128916023	128916148	MACS_peak_522	56.69
chr3	130770312	130770432	MACS_peak_523	57.36
chr3	137377529	137377669	MACS_peak_524	50.16
chr3	139850790	139850920	MACS_peak_525	54.83
chr3	141197718	141197841	MACS_peak_526	57.94
chr3	155379842	155379981	MACS_peak_527	52.43
chr3	190510583	190510727	MACS_peak_528	52.24
chr3	193203222	193203354	MACS_peak_529	54.97
chr3	196625587	196625783	MACS_peak_530	61.51
chr4	315211	315350	MACS_peak_531	52.73
chr4	1421135	1421606	MACS_peak_532	52.48
chr4	9345120	9345243	MACS_peak_533	62.79
chr4	13514952	13515116	MACS_peak_534	53.03
chr4	17111146	17111326	MACS_peak_535	53.14
chr4	18838758	18838849	MACS_peak_536	67.39
chr4	22174262	22174455	MACS_peak_537	66.51
chr4	30196323	30196468	MACS_peak_538	50.22
chr4	31781363	31781509	MACS_peak_539	54.71
chr4	37681094	37681271	MACS_peak_540	50.72
chr4	37793210	37793377	MACS_peak_541	56.26
chr4	39455075	39455198	MACS_peak_542	55.37
chr4	48134118	48134221	MACS_peak_543	62.88
chr4	49104297	49105151	MACS_peak_544	50.1
chr4	49111472	49112207	MACS_peak_545	58.38
chr4	49139028	49139412	MACS_peak_546	64.45
chr4	49144090	49144599	MACS_peak_547	50.91
chr4	49149218	49149865	MACS_peak_548	106.52
chr4	49151116	49151903	MACS_peak_549	83.47
chr4	49154822	49155867	MACS_peak_550	58.71
chr4	49280690	49280961	MACS_peak_551	62.21
chr4	49298010	49298444	MACS_peak_552	60.21
chr4	62267543	62267694	MACS_peak_553	50.72
chr4	62508985	62509131	MACS_peak_554	60.29
chr4	66743465	66743614	MACS_peak_555	51.18

chr4	73520826	73520953	MACS_peak_556	56.19
chr4	99297840	99297979	MACS_peak_557	53.35
chr4	100380816	100380940	MACS_peak_558	56.95
chr4	101636005	101636175	MACS_peak_559	56.32
chr4	118790791	118790932	MACS_peak_560	52.9
chr4	119326874	119327032	MACS_peak_561	58.89
chr4	122062330	122062484	MACS_peak_562	59.79
chr4	141109841	141109971	MACS_peak_563	65.81
chr4	148326671	148326811	MACS_peak_564	53.12
chr4	156935018	156935168	MACS_peak_565	70.91
chr4	162987462	162987585	MACS_peak_566	57.2
chr4	167224357	167224503	MACS_peak_567	51.81
chr4	177532851	177533001	MACS_peak_568	50.97
chr4	182234238	182234396	MACS_peak_569	50.72
chr4	183277044	183277189	MACS_peak_570	52.03
chr4	183390437	183390577	MACS_peak_571	52.21
chr4	186988536	186988670	MACS_peak_572	54.5
chr4	190865091	190865230	MACS_peak_573	51.44
chr4	191023045	191023231	MACS_peak_574	76.84
chr4_gl000193_random	221	343	MACS_peak_575	64.68
chr4_gl000193_random	163571	163694	MACS_peak_576	57.94
chr4_gl000194_random	147424	147669	MACS_peak_577	51.88
chr4_gl000194_random	164193	164299	MACS_peak_578	53.3
chr4_gl000194_random	167349	167674	MACS_peak_579	50.96
chr5	734522	734640	MACS_peak_580	56.98
chr5	14271974	14272133	MACS_peak_581	54.1
chr5	16271118	16271246	MACS_peak_582	54.5
chr5	21496788	21497008	MACS_peak_583	57.51
chr5	24127798	24127907	MACS_peak_584	53.24
chr5	30895997	30896149	MACS_peak_585	50.72
chr5	31400021	31400157	MACS_peak_586	64.2
chr5	34390231	34390341	MACS_peak_587	50.19
chr5	35477846	35477993	MACS_peak_588	50.39
chr5	45600002	45600156	MACS_peak_589	56.77
chr5	49405642	49406112	MACS_peak_590	55.58
chr5	49409303	49409853	MACS_peak_591	63.21
chr5	49413381	49414964	MACS_peak_592	54.73
chr5	49420629	49421438	MACS_peak_593	54.72
chr5	49423062	49424314	MACS_peak_594	82.34
chr5	49428790	49429763	MACS_peak_595	76.56
chr5	49429807	49430530	MACS_peak_596	71.33
chr5	49433186	49433830	MACS_peak_597	60.84
chr5	49435590	49436210	MACS_peak_598	53.31

chr5	49436395	49437600	MACS_peak_599	70.62
chr5	49438183	49440474	MACS_peak_600	119.95
chr5	57470872	57471019	MACS_peak_601	51.6
chr5	57672610	57672777	MACS_peak_602	65.52
chr5	61238124	61238262	MACS_peak_603	51.47
chr5	63143495	63143639	MACS_peak_604	51.03
chr5	65679687	65679836	MACS_peak_605	51.18
chr5	66815037	66815189	MACS_peak_606	60.25
chr5	69111038	69111212	MACS_peak_607	60.83
chr5	69578199	69578319	MACS_peak_608	52.12
chr5	69835902	69836028	MACS_peak_609	60.15
chr5	70289030	70289171	MACS_peak_610	52.6
chr5	70334968	70335090	MACS_peak_611	57.46
chr5	70469345	70469500	MACS_peak_612	50.47
chr5	70555630	70555782	MACS_peak_613	50.61
chr5	78923675	78923826	MACS_peak_614	50.77
chr5	93396865	93397001	MACS_peak_615	54.03
chr5	93596298	93596472	MACS_peak_616	64.48
chr5	95180161	95180306	MACS_peak_617	52.03
chr5	96148985	96149120	MACS_peak_618	50.73
chr5	99711978	99712113	MACS_peak_619	50.72
chr5	99826434	99826602	MACS_peak_620	50.72
chr5	122029674	122029798	MACS_peak_621	53.38
chr5	123991365	123991514	MACS_peak_622	60.95
chr5	126583023	126583169	MACS_peak_623	50.72
chr5	126933873	126934031	MACS_peak_624	50.72
chr5	127059088	127059233	MACS_peak_625	52.03
chr5	130678498	130678613	MACS_peak_626	59.34
chr5	131683469	131683616	MACS_peak_627	50.99
chr5	131822606	131822817	MACS_peak_628	106.72
chr5	132434376	132434521	MACS_peak_629	51.42
chr5	132938370	132938515	MACS_peak_630	52.03
chr5	143539489	143539626	MACS_peak_631	50.83
chr5	145534916	145535068	MACS_peak_632	50.72
chr5	149582463	149582609	MACS_peak_633	51.81
chr5	161272993	161273173	MACS_peak_634	63.75
chr5	161780243	161780395	MACS_peak_635	56.91
chr5	164500759	164500888	MACS_peak_636	53.28
chr5	169463339	169463463	MACS_peak_637	56.95
chr5	173992404	173992548	MACS_peak_638	52.24
chr5	177444115	177444253	MACS_peak_639	51.91
chr5	180836679	180836813	MACS_peak_640	54.5
chr6	2855319	2855458	MACS_peak_641	53.35

chr6	17311983	17312120	MACS_peak_642	53.8
chr6	21452224	21452375	MACS_peak_643	50.77
chr6	42298070	42298214	MACS_peak_644	52.24
chr6	51003686	51003824	MACS_peak_645	53.57
chr6	57248055	57248316	MACS_peak_646	55.28
chr6	58776036	58780166	MACS_peak_647	745.2
chr6	63363072	63363234	MACS_peak_648	59.68
chr6	71769955	71770064	MACS_peak_649	61.06
chr6	79685969	79686118	MACS_peak_650	51.18
chr6	85551134	85551266	MACS_peak_651	51.99
chr6	86549123	86549260	MACS_peak_652	53.8
chr6	103653103	103653179	MACS_peak_653	98.71
chr6	107459361	107459486	MACS_peak_654	56.69
chr6	120507395	120507483	MACS_peak_655	69.95
chr6	121487255	121487372	MACS_peak_656	58.17
chr6	131589063	131589203	MACS_peak_657	53.12
chr6	138309477	138309596	MACS_peak_658	58.25
chr6	144360436	144360603	MACS_peak_659	56.94
chr6	158838679	158838819	MACS_peak_660	51.31
chr6	171051568	171051715	MACS_peak_661	50.24
chr6_mann_hap4	1348551	1348681	MACS_peak_662	55.45
chr6_mcf_hap5	991108	991216	MACS_peak_663	84.12
chr7	74746	74950	MACS_peak_664	58.92
chr7	5065748	5065870	MACS_peak_665	51.73
chr7	9325693	9325820	MACS_peak_666	55.57
chr7	10785018	10785186	MACS_peak_667	59.32
chr7	17029259	17029404	MACS_peak_668	50.22
chr7	17203214	17203349	MACS_peak_669	50.18
chr7	20165435	20165571	MACS_peak_670	54.03
chr7	21472031	21472170	MACS_peak_671	50.38
chr7	24455183	24455305	MACS_peak_672	51.73
chr7	28567563	28567677	MACS_peak_673	51.35
chr7	30131442	30131582	MACS_peak_674	52.51
chr7	32037244	32037391	MACS_peak_675	61.43
chr7	34340086	34340224	MACS_peak_676	53.57
chr7	39078809	39078949	MACS_peak_677	56.74
chr7	49018529	49018646	MACS_peak_678	56.95
chr7	49999505	49999712	MACS_peak_679	65.11
chr7	51864366	51864491	MACS_peak_680	53.13
chr7	61111513	61111709	MACS_peak_681	72.8
chr7	61112179	61112355	MACS_peak_682	54.56
chr7	61152324	61152463	MACS_peak_683	51.9
chr7	61191470	61191681	MACS_peak_684	56.79

chr7	61370383	61370506	MACS_peak_685	56.71
chr7	61443176	61443246	MACS_peak_686	80.35
chr7	61638851	61639181	MACS_peak_687	57.81
chr7	61765170	61765443	MACS_peak_688	70.69
chr7	61967157	61967752	MACS_peak_689	138.16
chr7	61967781	61968148	MACS_peak_690	144.57
chr7	61968200	61969787	MACS_peak_691	104.28
chr7	61969797	61970609	MACS_peak_692	219.25
chr7	61972899	61973230	MACS_peak_693	67.76
chr7	61975370	61975625	MACS_peak_694	51.15
chr7	61976024	61976837	MACS_peak_695	102.79
chr7	62043636	62043724	MACS_peak_696	65.21
chr7	65238701	65238831	MACS_peak_697	55.45
chr7	66635337	66635445	MACS_peak_698	50.73
chr7	70738692	70738837	MACS_peak_699	52.03
chr7	70758328	70758469	MACS_peak_700	52.9
chr7	70904857	70904996	MACS_peak_701	53.35
chr7	74342286	74342426	MACS_peak_702	53.12
chr7	79675281	79675415	MACS_peak_703	54.5
chr7	88981655	88981836	MACS_peak_704	63.53
chr7	91109918	91110041	MACS_peak_705	64.71
chr7	100638332	100638451	MACS_peak_706	59.24
chr7	108239574	108239744	MACS_peak_707	50.51
chr7	111845173	111845282	MACS_peak_708	58.02
chr7	114824366	114824528	MACS_peak_709	58.16
chr7	114872270	114872400	MACS_peak_710	50.28
chr7	121650807	121650929	MACS_peak_711	51.73
chr7	121917663	121917783	MACS_peak_712	61.29
chr7	122806753	122806894	MACS_peak_713	51.39
chr7	126225858	126226006	MACS_peak_714	51.39
chr7	134799049	134799170	MACS_peak_715	53.05
chr7	134831885	134831984	MACS_peak_716	54.83
chr7	137033965	137034106	MACS_peak_717	50.22
chr7	140882976	140883119	MACS_peak_718	83.61
chr7	143775245	143775383	MACS_peak_719	50.89
chr7	144928021	144928144	MACS_peak_720	57.94
chr7	150568036	150568183	MACS_peak_721	51.3
chr7	152768968	152769117	MACS_peak_722	51.18
chr7	152841739	152841847	MACS_peak_723	61.35
chr7	158652521	158652652	MACS_peak_724	55.21
chr8	8912895	8913027	MACS_peak_725	54.97
chr8	12338297	12338429	MACS_peak_726	54.97
chr8	23934009	23934093	MACS_peak_727	74.59

chr8	27800702	27800843	MACS_peak_728	52.9
chr8	29936324	29936461	MACS_peak_729	53.8
chr8	38355335	38355465	MACS_peak_730	55.45
chr8	43793072	43794310	MACS_peak_731	193.84
chr8	43794596	43794976	MACS_peak_732	53.81
chr8	43798822	43798998	MACS_peak_733	52.77
chr8	43820864	43821269	MACS_peak_734	84.63
chr8	43822046	43823927	MACS_peak_735	155.84
chr8	43823932	43825510	MACS_peak_736	60.3
chr8	43826570	43827611	MACS_peak_737	61.21
chr8	43827896	43828750	MACS_peak_738	82.69
chr8	43829741	43830209	MACS_peak_739	74.56
chr8	43832133	43834065	MACS_peak_740	159.08
chr8	43834107	43835826	MACS_peak_741	177.57
chr8	43836979	43838761	MACS_peak_742	228.68
chr8	46838888	46839443	MACS_peak_743	61.62
chr8	46839589	46840793	MACS_peak_744	71.35
chr8	46840904	46841850	MACS_peak_745	73.49
chr8	46841908	46842670	MACS_peak_746	117.96
chr8	46843922	46844542	MACS_peak_747	69.68
chr8	46844572	46845442	MACS_peak_748	123.19
chr8	46846702	46846979	MACS_peak_749	64.52
chr8	46848195	46848810	MACS_peak_750	71
chr8	46848903	46851059	MACS_peak_751	227.63
chr8	46851295	46851604	MACS_peak_752	63.1
chr8	46852019	46854685	MACS_peak_753	320.1
chr8	46854778	46856547	MACS_peak_754	197.13
chr8	48761824	48761940	MACS_peak_755	58.13
chr8	51638825	51638970	MACS_peak_756	52.03
chr8	53004624	53004772	MACS_peak_757	51.39
chr8	53996462	53996606	MACS_peak_758	52.24
chr8	56691072	56691211	MACS_peak_759	53.35
chr8	63988052	63988192	MACS_peak_760	50.45
chr8	64443306	64443433	MACS_peak_761	56.19
chr8	68519394	68519483	MACS_peak_762	70.36
chr8	76008789	76008907	MACS_peak_763	55.22
chr8	83422072	83422250	MACS_peak_764	55.02
chr8	85314837	85314982	MACS_peak_765	52.03
chr8	85383947	85384083	MACS_peak_766	53.11
chr8	86729581	86730223	MACS_peak_767	66.85
chr8	86832011	86832308	MACS_peak_768	59.94
chr8	87031374	87031571	MACS_peak_769	69.68
chr8	87331750	87331894	MACS_peak_770	51.94

chr8	89313614	89313759	MACS_peak_771	52.03
chr8	89700427	89700553	MACS_peak_772	53.16
chr8	95688109	95688311	MACS_peak_773	56.36
chr8	99980648	99980751	MACS_peak_774	59.25
chr8	103568392	103568513	MACS_peak_775	57.1
chr8	104128349	104128497	MACS_peak_776	59.14
chr8	106416220	106416304	MACS_peak_777	60.86
chr8	109055059	109055196	MACS_peak_778	74.52
chr8	112176700	112176852	MACS_peak_779	70
chr8	112441470	112441621	MACS_peak_780	50.77
chr8	117679268	117679453	MACS_peak_781	53.27
chr8	121808718	121808873	MACS_peak_782	54.38
chr8	125551772	125551917	MACS_peak_783	52.03
chr8	125877518	125877716	MACS_peak_784	67.8
chr8	127591626	127591758	MACS_peak_785	52.28
chr8	129239040	129239203	MACS_peak_786	59.68
chr8	132449108	132449320	MACS_peak_787	53.85
chr8	135012515	135012633	MACS_peak_788	58.52
chr9	10233	10416	MACS_peak_789	130.81
chr9	24355	24486	MACS_peak_790	65.54
chr9	4942192	4942316	MACS_peak_791	57.67
chr9	6647558	6647702	MACS_peak_792	58.8
chr9	7358155	7358307	MACS_peak_793	50.72
chr9	9841707	9841844	MACS_peak_794	50.26
chr9	10864736	10864934	MACS_peak_795	55.12
chr9	15103731	15103874	MACS_peak_796	52.7
chr9	22174331	22174467	MACS_peak_797	54.17
chr9	23603919	23604061	MACS_peak_798	50.72
chr9	36568761	36568912	MACS_peak_799	50.77
chr9	39183107	39183261	MACS_peak_800	50.69
chr9	42031647	42031772	MACS_peak_801	50.98
chr9	42720389	42720571	MACS_peak_802	53.96
chr9	42912855	42912958	MACS_peak_803	54.05
chr9	45948796	45948946	MACS_peak_804	50.97
chr9	46679645	46679764	MACS_peak_805	54.96
chr9	47235157	47235304	MACS_peak_806	51.6
chr9	66293646	66293804	MACS_peak_807	53.24
chr9	66802083	66802476	MACS_peak_808	53.85
chr9	66833684	66834121	MACS_peak_809	78.98
chr9	66841563	66841735	MACS_peak_810	68.02
chr9	66857862	66858015	MACS_peak_811	51.28
chr9	66970955	66971521	MACS_peak_812	117.35
chr9	67320278	67320926	MACS_peak_813	135.68

chr9	69195642	69195798	MACS_peak_814	51.86
chr9	69533225	69533388	MACS_peak_815	50.93
chr9	69710833	69711378	MACS_peak_816	93.87
chr9	69993197	69993318	MACS_peak_817	52.53
chr9	70076689	70076800	MACS_peak_818	64.24
chr9	74335592	74335744	MACS_peak_819	53.32
chr9	75113398	75113520	MACS_peak_820	56.84
chr9	76936773	76936927	MACS_peak_821	51.76
chr9	78208234	78208375	MACS_peak_822	52.6
chr9	80611797	80611926	MACS_peak_823	51.59
chr9	81338824	81338975	MACS_peak_824	54.7
chr9	85283202	85283337	MACS_peak_825	51.29
chr9	93502599	93502742	MACS_peak_826	52.46
chr9	101020101	101020225	MACS_peak_827	51.23
chr9	102589025	102589164	MACS_peak_828	53.35
chr9	105359508	105359656	MACS_peak_829	50.18
chr9	109012269	109012400	MACS_peak_830	65.54
chr9	111206348	111206487	MACS_peak_831	51.24
chr9	111731279	111731347	MACS_peak_832	52.21
chr9	115458086	115458169	MACS_peak_833	139.65
chr9	118553570	118553717	MACS_peak_834	50.99
chr9	123843520	123843675	MACS_peak_835	80.07
chr9	129738452	129738598	MACS_peak_836	55.28
chr9	136955818	136955945	MACS_peak_837	56.19
chr9	140444191	140444336	MACS_peak_838	55.52
chr9_gl000198_random	9	220	MACS_peak_839	61.7
chr9_gl000199_random	34315	34714	MACS_peak_840	75.9
chr9_gl000199_random	37206	38766	MACS_peak_841	166.93
chr9_gl000199_random	39990	41788	MACS_peak_842	126.48
chr9_gl000199_random	44046	44881	MACS_peak_843	178.53
chr9_gl000199_random	49446	50501	MACS_peak_844	74.17
chr9_gl000199_random	50545	52869	MACS_peak_845	70.14
chr9_gl000199_random	52952	53637	MACS_peak_846	79.73
chr9_gl000199_random	53694	56065	MACS_peak_847	184.62
chr9_gl000199_random	57042	57768	MACS_peak_848	77.51
chr9_gl000199_random	58081	58895	MACS_peak_849	109.21
chr9_gl000199_random	59318	59976	MACS_peak_850	89.48
chr9_gl000199_random	64723	67974	MACS_peak_851	322.35
chr9_gl000199_random	67986	68925	MACS_peak_852	73.49
chr9_gl000199_random	68980	69847	MACS_peak_853	92.23
chr9_gl000199_random	71398	73244	MACS_peak_854	197.55
chr9_gl000199_random	73294	75949	MACS_peak_855	148.96
chr9_gl000199_random	76415	79668	MACS_peak_856	273.95

chr9_gl000199_random	79693	81983	MACS_peak_857	96.09
chr9_gl000199_random	82762	83556	MACS_peak_858	115.11
chr9_gl000199_random	83995	85440	MACS_peak_859	210.34
chr9_gl000199_random	86193	86570	MACS_peak_860	77.64
chr9_gl000199_random	88892	89996	MACS_peak_861	99.56
chr9_gl000199_random	90039	93249	MACS_peak_862	305.4
chr9_gl000199_random	93280	94161	MACS_peak_863	85.17
chr9_gl000199_random	94208	94800	MACS_peak_864	51.91
chr9_gl000199_random	95153	95433	MACS_peak_865	74.64
chr9_gl000199_random	95460	96935	MACS_peak_866	105.96
chr9_gl000199_random	98688	100760	MACS_peak_867	186.74
chr9_gl000199_random	100874	102301	MACS_peak_868	144.09
chr9_gl000199_random	102325	103491	MACS_peak_869	80.71
chr9_gl000199_random	103822	111014	MACS_peak_870	457.81
chr9_gl000199_random	115902	117672	MACS_peak_871	206.92
chr9_gl000199_random	117729	121494	MACS_peak_872	210.56
chr9_gl000199_random	121514	123367	MACS_peak_873	146.49
chr9_gl000199_random	124083	125340	MACS_peak_874	96.71
chr9_gl000199_random	129128	130171	MACS_peak_875	68.6
chr9_gl000199_random	130182	131118	MACS_peak_876	94.54
chr9_gl000199_random	133347	136323	MACS_peak_877	242.64
chr9_gl000199_random	136327	137563	MACS_peak_878	102.82
chr9_gl000199_random	137693	138598	MACS_peak_879	83.91
chr9_gl000199_random	139003	139338	MACS_peak_880	59.78
chr9_gl000199_random	140081	141811	MACS_peak_881	132.05
chr9_gl000199_random	144596	147725	MACS_peak_882	230.96
chr9_gl000199_random	147916	148224	MACS_peak_883	57.53
chr9_gl000199_random	148244	148801	MACS_peak_884	89.85
chr9_gl000199_random	149072	150717	MACS_peak_885	144.25
chr9_gl000199_random	150752	152147	MACS_peak_886	84.49
chr9_gl000199_random	161760	162658	MACS_peak_887	80.46
chr9_gl000199_random	162676	163505	MACS_peak_888	78.42
chr9_gl000199_random	163507	165205	MACS_peak_889	116.6
chrUn_gl000212	138964	139101	MACS_peak_890	56.35
chrUn_gl000214	33297	33470	MACS_peak_891	51.22
chrUn_gl000214	64088	65798	MACS_peak_892	132.48
chrUn_gl000214	106942	107174	MACS_peak_893	63
chrUn_gl000214	121652	122485	MACS_peak_894	53.52
chrUn_gl000214	132511	134097	MACS_peak_895	104.46
chrUn_gl000216	75602	76210	MACS_peak_896	84.12
chrUn_gl000216	135841	136062	MACS_peak_897	77.2
chrUn_gl000216	145038	145602	MACS_peak_898	76.62
chrUn_gl000216	151115	151523	MACS_peak_899	54.61

chrUn_gl000216	154628	155373	MACS_peak_900	53.31
chrUn_gl000216	158688	159023	MACS_peak_901	58.97
chrUn_gl000216	159970	160378	MACS_peak_902	102.03
chrUn_gl000216	163422	163775	MACS_peak_903	72.8
chrUn_gl000217	112440	112684	MACS_peak_904	52.71
chrUn_gl000217	166135	166313	MACS_peak_905	56.94
chrUn_gl000218	129643	129906	MACS_peak_906	54.74
chrUn_gl000221	36358	36515	MACS_peak_907	57.74
chrUn_gl000224	13466	14649	MACS_peak_908	55.12
chrUn_gl000225	14369	14783	MACS_peak_909	52.28
chrUn_gl000225	15028	15343	MACS_peak_910	52.64
chrUn_gl000225	19573	19740	MACS_peak_911	61.35
chrUn_gl000225	21368	22032	MACS_peak_912	52.9
chrUn_gl000225	25328	25812	MACS_peak_913	71.53
chrUn_gl000225	49763	50088	MACS_peak_914	54.12
chrUn_gl000225	50770	51280	MACS_peak_915	55.97
chrUn_gl000225	67244	67686	MACS_peak_916	69.46
chrUn_gl000225	92244	92957	MACS_peak_917	60.73
chrUn_gl000225	101538	101914	MACS_peak_918	76.52
chrUn_gl000225	103233	103362	MACS_peak_919	51.1
chrUn_gl000225	110203	110636	MACS_peak_920	84.03
chrUn_gl000225	111350	111621	MACS_peak_921	77.63
chrUn_gl000225	116685	117113	MACS_peak_922	60.27
chrUn_gl000225	117652	118266	MACS_peak_923	70.3
chrUn_gl000225	125109	125397	MACS_peak_924	50.02
chrUn_gl000225	129984	130614	MACS_peak_925	51.11
chrUn_gl000226	0	293	MACS_peak_926	76.02
chrUn_gl000228	117149	118879	MACS_peak_927	144.83
chrX	16355665	16355800	MACS_peak_928	75.11
chrX	38403869	38404012	MACS_peak_929	52.46
chrX	51364364	51364535	MACS_peak_930	77.18
chrX	55455751	55455902	MACS_peak_931	50.77
chrX	58562429	58562858	MACS_peak_932	51.2
chrX	58564701	58582004	MACS_peak_933	120.35
chrX	61682030	61683963	MACS_peak_934	190.15
chrX	61683970	61685607	MACS_peak_935	132.96
chrX	61686469	61687135	MACS_peak_936	70.13
chrX	61689024	61690487	MACS_peak_937	108.89
chrX	61690769	61693105	MACS_peak_938	53.45
chrX	61693129	61694821	MACS_peak_939	129.39
chrX	61694883	61696311	MACS_peak_940	104.28
chrX	61696360	61697102	MACS_peak_941	133.76
chrX	61697115	61697615	MACS_peak_942	70.25

chrX	61697622	61698843	MACS_peak_943	170.73
chrX	61698867	61700239	MACS_peak_944	147.85
chrX	61700300	61700869	MACS_peak_945	96.86
chrX	61700920	61701163	MACS_peak_946	55.37
chrX	61701736	61702621	MACS_peak_947	60.77
chrX	61702682	61703142	MACS_peak_948	58.58
chrX	61703529	61707158	MACS_peak_949	400.29
chrX	61707239	61707839	MACS_peak_950	69.73
chrX	61708738	61709980	MACS_peak_951	139.16
chrX	61709981	61710846	MACS_peak_952	64.58
chrX	61710893	61712928	MACS_peak_953	168.02
chrX	61713558	61714303	MACS_peak_954	115.53
chrX	61714375	61715631	MACS_peak_955	99.02
chrX	61716583	61716884	MACS_peak_956	58.25
chrX	61717716	61718548	MACS_peak_957	54.27
chrX	61718596	61720181	MACS_peak_958	99.45
chrX	61720182	61721060	MACS_peak_959	114.18
chrX	61722617	61723243	MACS_peak_960	115.53
chrX	61774234	61774401	MACS_peak_961	56.3
chrX	61858992	61859084	MACS_peak_962	50.46
chrX	73618210	73618371	MACS_peak_963	89.07
chrX	75543460	75543588	MACS_peak_964	64.97
chrX	75637907	75638052	MACS_peak_965	52.03
chrX	79352781	79352915	MACS_peak_966	64.73
chrX	91078601	91078714	MACS_peak_967	59.9
chrX	91227591	91227738	MACS_peak_968	51.6
chrX	93742938	93743083	MACS_peak_969	52.03
chrX	101833112	101833240	MACS_peak_970	55.94
chrX	103606445	103606530	MACS_peak_971	65.11
chrX	108297715	108297788	MACS_peak_972	72.01
chrX	109353034	109353184	MACS_peak_973	50.97
chrX	119477522	119477655	MACS_peak_974	51.76
chrX	122209556	122209701	MACS_peak_975	52.03
chrX	151585407	151585551	MACS_peak_976	52.24
chrY	8333466	8333590	MACS_peak_977	56.95
chrY	9994655	9994790	MACS_peak_978	51.29
chrY	13113536	13113624	MACS_peak_979	54.07
chrY	13413058	13413193	MACS_peak_980	52.3
chrY	13468673	13469203	MACS_peak_981	61.29
chrY	13474208	13474608	MACS_peak_982	50.4
chrY	13482020	13482203	MACS_peak_983	59.4
chrY	13485431	13485832	MACS_peak_984	98.04
chrY	13547997	13548148	MACS_peak_985	60.48

chrY	13842554	13842838	MACS_peak_986	62.44
chrY	13845529	13845593	MACS_peak_987	53.38
chrY	26170430	26170559	MACS_peak_988	52.21
chrY	58996883	58997192	MACS_peak_989	53.76

Appendix VI: List of Proteins containing [ATMPVSAG] motif identified in phage display screen

Uniprot ID	*Protein ID*	*Name*	*Pattern*	*Sequence*
Q9NQ94	A1CF	APOBEC1 complementation factor	A-T-X-P-V-S-A	ATaPVSA
O15178	BRAC	Brachyury protein	T-X-P-V-S-A	ThPVSA
Q5VW32	BROX	BRO1 domain-containing protein BROX	A-T-X-P-V-S	ATaPVS
Q16602	CALRL	Calcitonin gene-related peptide type 1 receptor	A-T-X-P-V-S	ATnPVS
Q9ULU8	CAPS1	Calcium-dependent secretion activator 1	M-P-V-S-X-G	MPVSkG
Q6ZP82	CC141	Coiled-coil domain-containing protein 141	A-X-M-P-V-S-A	AlMPVSA
Q4KMG0	CDON	Cell adhesion molecule-related/down-regulated by oncogenes	T-X-P-V-S-A	TaPVSA
Q8N9S7	CHDC2	Calponin homology domain-containing protein 2	T-M-P-V-S	TMPVS
Q9BUW7	CI016	UPF0184 protein C9orf16	M-P-V-X-A-G	MPVeAG
Q03692	COAA1	Collagen alpha-1(X) chain	M-P-V-S-A	MPVSA
P11511	CP19A	Aromatase	A-T-M-P-V	ATMPV
P10632	CP2C8	Cytochrome P450 2C8	M-P-X-S-A-G	MPfSAG
P11712	CP2C9	Cytochrome P450 2C9	M-P-X-S-A-G	MPfSAG
P33260	CP2CI	Cytochrome P450 2C18	M-P-X-S-A-G	MPfSAG
P33261	CP2CJ	Cytochrome P450 2C19	M-P-X-S-A-G	MPfSAG
P24903	CP2F1	Cytochrome P450 2F1	M-P-X-S-A-G	MPfSAG
P13584	CP4B1	Cytochrome P450 4B1	M-P-X-S-A-G	MPfSAG
Q9H7T0	CTSRB	Cation channel sperm-associated protein subunit beta	A-T-M-P-X-S-A	ATMPeSA
Q07687	DLX2	Homeobox protein DLX-2	P-V-S-A-G	PVSAG
Q9N2J8	ENH3	HERV-H_2q241 provirus ancestral Env polyprotein	A-T-X-P-V-S	ATsPVS
Q9Y4F4	F179B	Protein FAM179B	M-P-X-S-A-G	MPsSAG
Q8NE31	FA13C	Protein FAM13C	A-T-M-P-V	ATMPV
Q8N5W8	FA24B	Protein FAM24B	M-P-V-X-A-G	MPViAG
Q12805	FBLN3	EGF-containing fibulin-like extracellular matrix protein 1	T-X-P-V-S-A	TsPVSA
Q68DA7	FMN1	Formin-1	P-V-S-A-G	PVSAG
Q8WXD5	GEMI6	Gem-associated protein 6	T-X-P-V-S-A	TdPVSA
Q53GS7	GLE1	Nucleoporin GLE1	A-T-X-P-V-S	ATiPVS
Q3T8J9	GON4L	GON-4-like protein	M-P-V-S-A-G	MPVSAG
P56524	HDAC4	Histone deacetylase 4	A-T-X-P-V-S	ATfPVS
Q86VS8	HOOK3	Protein Hook homolog 3	P-V-S-A-G	PVSAG
Q9BQS6	HSPB9	Heat shock protein beta-9	A-T-M-P-V	ATMPV
Q7Z6Z7	HUWE1	E3 ubiquitin-protein ligase HUWE1	P-V-S-A-G	PVSAG
Q17R60	IMPG1	Interphotoreceptor matrix proteoglycan 1	T-X-P-V-S-A	TtPVSA

P35568	IRS1	Insulin receptor substrate 1	A-T-X-P-V-S	ATsPVS
Q14571	ITPR2	Inositol 1,4,5-trisphosphate receptor type 2	M-P-V-X-A-G	MPVnAG
P52333	JAK3	Tyrosine-protein kinase JAK3	T-M-P-X-S-A	TMPiSA
Q15652	JHD2C	Probable JmjC domain-containing histone demethylation protein 2C	M-P-V-X-A-G	MPVnAG
P17275	JUNB	Transcription factor jun-B	A-T-X-P-V-S	ATpPVS
Q5XKE5	K2C79	Keratin, type II cytoskeletal 79	P-V-S-A-G	PVSAG
P43632	KI2S4	Killer cell immunoglobulin-like receptor 2DS4	M-P-V-X-A-G	MPVlAG
P43630	KI3L2	Killer cell immunoglobulin-like receptor 3DL2	M-P-V-X-A-G	MPVlAG
P06858	LIPL	Lipoprotein lipase	P-V-S-A-G	PVSAG
Q6ZMQ8	LMTK1	Serine/threonine-protein kinase LMTK1	P-V-S-A-G	PVSAG
Q93052	LPP	Lipoma-preferred partner	T-X-P-V-S-A	TdPVSA
Q96NW7	LRRC7	Leucine-rich repeat-containing protein 7	T-X-P-V-S-A	TrPVSA
Q3C1V0	M4A18	Membrane-spanning 4-domains subfamily A member 18	A-T-X-P-V-S-A	ATgPVSA
Q8TC05	MDM1	Nuclear protein MDM1	A-T-X-P-V-S	ATlPVS
O43451	MGA	Maltase-glucoamylase, intestinal	T-X-P-V-S-A	TtPVSA
Q5HYA8	MKS3	Meckelin	A-T-X-P-V-S	ATvPVS
Q99549	MPP8	M-phase phosphoprotein 8	M-P-V-S-A	MPVSA
Q8WXI7	MUC16	Mucin-16	T-M-P-X-S-A	TMPnSA
Q685J3	MUC17	Mucin-17	A-X-M-P-V-S	AsMPVS
Q6W4X9	MUC6	Mucin-6	A-T-X-P-V-S	ATtPVS
E9PAV3	NACAM	Nascent polypeptide-associated complex subunit alpha, muscle-specific form	A-X-M-P-V-S	AqMPVS
Q3BBV0	NBPF1	Neuroblastoma breakpoint family member 1	M-X-V-S-A-G	MvVSAG
Q9H094	NBPF3	Neuroblastoma breakpoint family member 3	M-X-V-S-A-G	MvVSAG
P0C2Y1	NBPF7	Putative neuroblastoma breakpoint family member 7	T-M-X-V-S-A-G	TMvVSAG
Q3BBV2	NBPF8	Putative neuroblastoma breakpoint family member 8	M-X-V-S-A-G	MvVSAG
Q3BBW0	NBPF9	Neuroblastoma breakpoint family member 9	M-X-V-S-A-G	MvVSAG
Q6P3W6	NBPFA	Neuroblastoma breakpoint family member 10	M-X-V-S-A-G	MvVSAG
Q86T75	NBPFB	Neuroblastoma breakpoint family member 11	M-X-V-S-A-G	MvVSAG
Q8N660	NBPFF	Neuroblastoma breakpoint family member 15	M-X-V-S-A-G	MvVSAG
Q5SXJ2	NBPFG	Neuroblastoma breakpoint family member 16	M-X-V-S-A-G	MvVSAG
Q3BBV1	NBPFK	Neuroblastoma breakpoint family member 20	M-X-V-S-A-G	MvVSAG
A6NDD8	NBPFL	Neuroblastoma breakpoint family member 21	M-X-V-S-A-G	MvVSAG

Q5RGN0	NBPFN	Putative neuroblastoma breakpoint family member 24	M-X-V-S-A-G	MvVSAG
Q13133	NR1H3	Oxysterols receptor LXR-alpha	M-P-X-S-A-G	MPhSAG
Q92621	NU205	Nuclear pore complex protein Nup205	M-P-V-X-A-G	MPVvAG
Q8NFH3	NUP43	Nucleoporin Nup43	T-M-P-V-S	TMPVS
Q9UKX7	NUP50	Nuclear pore complex protein Nup50	A-T-M-P-V	ATMPV
Q8N328	PGBD3	PiggyBac transposable element-derived protein 3	A-T-X-P-V-S	ATaPVS
Q15735	PI5PA	Phosphatidylinositol 4,5-bisphosphate 5-phosphatase A	M-P-X-S-A-G	MPaSAG
Q7Z443	PK1L3	Polycystic kidney disease protein 1-like 3	T-M-P-V-S	TMPVS
Q6ZV29	PLPL7	Patatin-like phospholipase domain-containing protein 7	A-X-M-P-V-S	AvMPVS
Q9ULL5	PRR12	Proline-rich protein 12	P-V-S-A-G	PVSAG
Q6WKZ4	RFIP1	Rab11 family-interacting protein 1	T-X-P-V-S-A	ThPVSA
Q92736	RYR2	Ryanodine receptor 2	M-P-X-S-A-G	MPlSAG
Q9BRG2	SH23A	SH2 domain-containing protein 3A	T-M-P-X-S-A	TMPiSA
Q9NR46	SHLB2	Endophilin-B2	A-T-M-P-V	ATMPV
Q4ZJI4	SL9B1	Sodium/hydrogen exchanger 9B1	T-M-X-V-S-A	TMcVSA
Q07890	SOS2	Son of sevenless homolog 2	T-X-P-V-S-A	TpPVSA
O15056	SYNJ2	Synaptojanin-2	P-V-S-A-G	PVSAG
Q5T011	SZT2	Protein SZT2	A-T-M-P-X-S-A-G	ATMPiSAG
O95359	TACC2	Transforming acidic coiled-coil-containing protein 2	A-X-M-P-V-S-A	AgMPVSA
Q12962	TAF10	Transcription initiation factor TFIID subunit 10	P-V-S-A-G	PVSAG
Q7Z7G0	TARSH	Target of Nesh-SH3	P-V-S-A-G	PVSAG
Q10587	TEF	Thyrotroph embryonic factor	T-M-X-V-S-A	TMaVSA
Q9UQP3	TENN	Tenascin-N	A-T-X-P-V-S	ATyPVS
Q53S58	TM177	Transmembrane protein 177	P-V-S-A-G	PVSAG
Q96B49	TOM6	Mitochondrial import receptor subunit TOM6 homolog	T-X-P-V-S-A	TvPVSA
P38435	VKGC	Vitamin K-dependent gamma-carboxylase	M-X-V-S-A-G	MaVSAG
Q5VTH9	WDR78	WD repeat-containing protein 78	A-T-M-P-V-S	ATMPVS
P59817	Z280A	Zinc finger protein 280A	A-X-M-P-V-S	AiMPVS
Q7Z3T8	ZFY16	Zinc finger FYVE domain-containing protein 16	A-T-X-P-V-S-A	ATlPVSA

Appendix VII: List of Proteins containing [NPQPSPMGS] motif identified in phage display screen

Uniprot ID	*Protein ID*	*Name*	*Pattern*	*Sequence*
O14734	ACOT8	Acyl-coenzyme A thioesterase 8	Q-P-S-P-M	QPSPM
P55196	AFAD	Afadin	Q-P-S-P-M	QPSPM
Q5T2P9	AGA10	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 10	N-P-X-P-S-P	NPpPSP
Q99490	AGAP2	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 2	P-X-P-S-P-M	PpPSPM
Q96P64	AGAP4	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 4	N-P-X-P-S-P	NPpPSP
A6NIR3	AGAP5	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 5	N-P-X-P-S-P	NPpPSP
Q5VW22	AGAP6	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 6	N-P-X-P-S-P	NPpPSP
Q5VUJ5	AGAP7	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 7	N-P-X-P-S-P	NPpPSP
Q5SRD3	AGAP8	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 8	N-P-X-P-S-P	NPpPSP
Q5VTM2	AGAP9	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 9	N-P-X-P-S-P	NPpPSP
Q7Z591	AKNA	AT-hook-containing transcription factor	N-P-Q-P-X-P	NPQPrP
Q86TB3	ALPK2	Alpha-protein kinase 2	N-X-Q-P-S-P	NkQPSP
Q86TB3	ALPK2	Alpha-protein kinase 2	N-P-Q-X-S-P	NPQlSP
Q12955	ANK3	Ankyrin-3	P-S-P-X-G-S	PSPqGS
P25054	APC	Adenomatous polyposis coli protein	P-X-P-M-G-S	PvPMGS
Q13794	APR	Phorbol-12-myristate-13-acetate-induced protein 1	N-X-Q-P-S-P	NaQPSP
A6NJG6	ARGFX	Arginine-fifty homeobox	N-P-Q-P-X-P	NPQPdP
O14497	ARI1A	AT-rich interactive domain-containing protein 1A	P-S-P-X-G-S	PSPvGS
Q8NFD5	ARI1B	AT-rich interactive domain-containing protein 1B	P-S-P-X-G-S	PSPvGS
Q9NZ32	ARP10	Actin-related protein 10	P-S-X-M-G-S	PSvMGS
Q76LX8	ATS13	A disintegrin and metalloproteinase with thrombospondin motifs 13	N-P-Q-P-X-P	NPQPcP
Q76LX8	ATS13	A disintegrin and metalloproteinase with thrombospondin motifs 13	P-S-P-X-G-S	PSPwGS
Q9UIF9	BAZ2A	Bromodomain adjacent to zinc finger domain protein 2A	P-Q-X-S-P-M	PQtSPM
Q8NFC6	BD1L1	Biorientation of chromosomes in cell division protein 1-like 1	N-P-Q-P-X-P	NPQPqP
Q9UHR4	BI2L1	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1	P-Q-X-S-P-M	PQaSPM
Q8IWQ3	BRSK2	Serine/threonine-protein kinase BRSK2	P-S-P-X-G-S	PSPrGS
Q9UPA5	BSN	Protein bassoon	P-X-P-S-P-M	PtPSPM

Q4AC94	C2CD3	C2 domain-containing protein 3	N-P-Q-P-X-P	NPQPiP
Q9H6R7	CB044	WD repeat-containing protein C2orf44	P-S-P-M-X-S	PSPMeS
Q76M96	CCD80	Coiled-coil domain-containing protein 80	P-S-P-M-X-S	PSPMwS
Q12834	CDC20	Cell division cycle protein 20 homolog	P-X-P-S-P-M	PaPSPM
O75956	CDKA2	Cyclin-dependent kinase 2-associated protein 2	P-S-P-X-G-S	PSPsGS
P19835	CEL	Bile salt-activated lipase	N-P-Q-P-X-P	NPQPhP
Q9BZC1	CELF4	CUGBP Elav-like family member 4	P-Q-P-X-P-M	PQPpPM
Q8N6W0	CELF5	CUGBP Elav-like family member 5	P-S-P-X-G-S	PSPvGS
Q92674	CENPI	Centromere protein I	P-S-X-M-G-S	PSqMGS
A6NGG3	CI092	Putative uncharacterized protein C9orf92	Q-P-S-P-X-G	QPSPtG
Q8N655	CJ012	Uncharacterized protein C10orf12	Q-X-S-P-M-G	QsSPMG
Q96F05	CK024	Uncharacterized protein C11orf24	Q-P-S-P-M	QPSPM
Q96M19	CL067	Transmembrane protein LINC00477	P-S-P-M-G	PSPMG
Q2M3C6	CO027	Transmembrane protein C15orf27	P-S-P-X-G-S	PSPaGS
P29400	CO4A5	Collagen alpha-5(IV) chain	Q-P-X-P-M-G	QPgPMG
Q99715	COCA1	Collagen alpha-1(XII) chain	N-P-X-P-S-P	NPsPSP
P39060	COIA1	Collagen alpha-1(XVIII) chain	P-Q-X-S-P-M	PQvSPM
Q96A83	COQA1	Collagen alpha-1(XXVI) chain	N-P-X-P-S-P	NPgPSP
P22792	CPN2	Carboxypeptidase N subunit 2	Q-P-X-P-M-G	QPcPMG
O94985	CSTN1	Calsyntenin-1	P-S-P-X-G-S	PSPsGS
P0CG12	CTF8A	Chromosome transmission fidelity protein 8 homolog isoform 2	P-X-P-S-P-M	PnPSPM
Q9Y6H8	CXA3	Gap junction alpha-3 protein	P-S-P-X-G-S	PSPvGS
Q6PJP8	DCR1A	DNA cross-link repair 1A protein	N-P-X-P-S-P	NPvPSP
Q8N1I0	DOCK4	Dedicator of cytokinesis protein 4	N-X-Q-P-S-P	NmQPSP
Q96BJ8	ELMO3	Engulfment and cell motility protein 3	S-P-M-G-S	SPMGS
Q14241	ELOA1	Transcription elongation factor B polypeptide 3	P-X-P-M-G-S	PyPMGS
Q09472	EP300	Histone acetyltransferase p300	N-P-Q-P-X-P	NPQPiP
Q96L91	EP400	E1A-binding protein p400	N-P-X-P-S-P	NPpPSP
Q6NXG1	ESRP1	Epithelial splicing regulatory protein 1	P-S-P-X-G-S	PSPpGS
Q86US8	EST1A	Telomerase-binding protein EST1A	P-S-P-X-G-S	PSPiGS
Q9BSJ8	ESYT1	Extended synaptotagmin-1	P-X-P-S-P-M	PsPSPM
Q8N612	F16A2	FTS and Hook-interacting protein	P-S-P-X-G-S	PSPfGS
Q32MH5	F214A	Protein FAM214A	N-P-Q-X-S-P	NPQfSP
A6ND36	FA83G	Protein FAM83G	N-P-Q-P-S-X-M	NPQPSqM
A6ND36	FA83G	Protein FAM83G	P-Q-P-S-X-M	PQPSqM
O14526	FCHO1	FCH domain only protein 1	P-S-P-X-G-S	PSPlGS
Q8WW38	FOG2	Zinc finger protein ZFPM2	Q-P-S-P-X-G	QPSPnG
Q9P0K8	FOXJ2	Forkhead box protein J2	P-X-P-S-P-M	PgPSPM
O43524	FOXO3	Forkhead box protein O3	Q-P-S-P-X-G	QPSPtG
Q9Y2L6	FRM4B	FERM domain-containing protein 4B	N-P-X-P-S-P	NPlPSP
Q5TBA9	FRY	Protein furry homolog	P-Q-P-X-P-M	PQPlPM

O94915	FRYL	Protein furry homolog-like	Q-P-S-P-M	QPSPM
Q14435	GALT3	Polypeptide N-acetylgalactosaminyltransferase 3	P-S-P-X-G-S	PSPyGS
Q9H3P7	GCP60	Golgi resident protein GCP60	P-S-P-X-G-S	PSPpGS
P10070	GLI2	Zinc finger protein GLI2	P-S-P-X-G-S	PSPlGS
P10071	GLI3	Transcriptional activator GLI3	Q-P-S-P-X-G	QPSPgG
Q96PE1	GP124	G-protein coupled receptor 124	P-X-P-S-P-M	PtPSPM
Q6ZVF9	GRIN3	G protein-regulated inducer of neurite outgrowth 3	N-P-X-P-S-P	NPtPSP
Q96CP6	GRM1A	GRAM domain-containing protein 1A	P-S-P-X-G-S	PSPvGS
Q7LDG7	GRP2	RAS guanyl-releasing protein 2	P-X-P-S-P-M	PsPSPM
P54840	GYS2	Glycogen [starch] synthase, liver	P-S-P-X-G-S	PSPsGS
P41235	HNF4A	Hepatocyte nuclear factor 4-alpha	P-Q-P-S-P-X-G	PQPSPpG
P41235	HNF4A	Hepatocyte nuclear factor 4-alpha	Q-P-S-P-X-G	QPSPpG
Q96MM6	HS12B	Heat shock 70 kDa protein 12B	P-S-P-X-G-S	PSPpGS
O43364	HXA2	Homeobox protein Hox-A2	P-S-P-X-G-S	PSPaGS
P14651	HXB3	Homeobox protein Hox-B3	P-S-P-X-G-S	PSPaGS
O14920	IKKB	Inhibitor of nuclear factor kappa-B kinase subunit beta	Q-X-S-P-M-G	QrSPMG
Q9C086	IN80B	INO80 complex subunit B	Q-P-S-P-M	QPSPM
Q15811	ITSN1	Intersectin-1	P-X-P-M-G-S	PvPMGS
Q8N9B5	JMY	Junction-mediating and -regulatory protein	N-P-X-P-S-P	NPlPSP
O60303	K0556	Uncharacterized protein KIAA0556	Q-P-S-P-X-G	QPSPtG
Q9P2N6	KANL3	KAT8 regulatory NSL complex subunit 3	S-P-M-G-S	SPMGS
Q92831	KAT2B	Histone acetyltransferase KAT2B	N-P-X-P-S-P	NPnPSP
Q14003	KCNC3	Potassium voltage-gated channel subfamily C member 3	P-S-P-X-G-S	PSPdGS
Q2KJY2	KI26B	Kinesin-like protein KIF26B	P-X-P-M-G-S	PrPMGS
Q96AC6	KIFC2	Kinesin-like protein KIFC2	P-S-P-X-G-S	PSPdGS
O14686	KMT2D	Histone-lysine N-methyltransferase 2D	P-X-P-S-P-M	PlPSPM
P18627	LAG3	Lymphocyte activation gene 3 protein	P-Q-X-S-P-M	PQvSPM
Q7Z429	LFG1	Protein lifeguard 1	Q-P-S-P-X-G	QPSPyG
Q14693	LPIN1	Phosphatidate phosphatase LPIN1	P-S-P-X-G-S	PSPsGS
O75197	LRP5	Low-density lipoprotein receptor-related protein 5	N-P-X-P-S-P	NPpPSP
O75581	LRP6	Low-density lipoprotein receptor-related protein 6	N-P-X-P-S-P	NPpPSP
Q13495	MAMD1	Mastermind-like domain-containing protein 1	N-P-X-P-S-P	NPgPSP
Q8IZL2	MAML2	Mastermind-like protein 2	P-Q-X-S-P-M	PQsSPM
P78559	MAP1A	Microtubule-associated protein 1A	P-X-P-M-G-S	PiPMGS
Q9UHV7	MED13	Mediator of RNA polymerase II transcription subunit 13	P-S-P-X-G-S	PSPvGS
Q8IVT2	MISP	Mitotic interactor and substrate of PLK1	Q-P-S-P-X-G	QPSPrG
P45983	MK08	Mitogen-activated protein kinase 8	Q-P-S-P-X-G	QPSPlG

P53779	MK10	Mitogen-activated protein kinase 10	Q-P-S-P-X-G	QPSPsG
A4FU01	MTMRB	Myotubularin-related protein 11	P-Q-P-S-X-M	PQPSfM
Q765P7	MTSSL	MTSS1-like protein	P-S-P-M-X-S	PSPMpS
Q8WXI7	MUC16	Mucin-16	P-S-P-M-X-S	PSPMtS
E2RYF6	MUC22	Mucin-22	P-S-P-X-G-S	PSPtGS
Q9NQX4	MYO5C	Unconventional myosin-Vc	P-S-P-X-G-S	PSPfGS
Q5VU43	MYOME	Myomegalin	P-S-P-M-G	PSPMG
Q14896	MYPC3	Myosin-binding protein C, cardiac-type	P-S-P-X-G-S	PSPkGS
Q8NFW9	MYRIP	Rab effector MyRIP	N-P-Q-P-X-P	NPQPqP
E9PAV3	NACAM	Nascent polypeptide-associated complex subunit alpha, muscle-specific form	P-S-P-X-G-S	PSPkGS
Q14686	NCOA6	Nuclear receptor coactivator 6	Q-P-S-X-M-G	QPSlMG
Q9Y618	NCOR2	Nuclear receptor corepressor 2	P-S-X-M-G-S	PSrMGS
O95644	NFAC1	Nuclear factor of activated T-cells, cytoplasmic 1	P-S-P-X-G-S	PSPhGS
Q0D2K0	NIPA4	Magnesium transporter NIPA4	N-P-X-P-S-P	NPpPSP
P20393	NR1D1	Nuclear receptor subfamily 1 group D member 1	P-S-P-X-G-S	PSPtGS
Q92570	NR4A3	Nuclear receptor subfamily 4 group A member 3	P-S-P-X-G-S	PSPpGS
Q9GZK4	OR2H1	Olfactory receptor 2H1	Q-X-S-P-M-G	QsSPMG
Q9BXW6	OSBL1	Oxysterol-binding protein-related protein 1	P-S-P-M-X-S	PSPMfS
Q96A73	P33MX	Putative monooxygenase p33MONOX	P-S-X-M-G-S	PSdMGS
Q9BTU6	P4K2A	Phosphatidylinositol 4-kinase type 2-alpha	P-S-P-X-G-S	PSPpGS
O00764	PDXK	Pyridoxal kinase	P-S-P-X-G-S	PSPqGS
Q8IXK0	PHC2	Polyhomeotic-like protein 2	N-X-Q-P-S-P	NiQPSP
Q86WI1	PKHL1	Fibrocystin-L	N-P-X-P-S-P	NPlPSP
Q9NQ66	PLCB1	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-1	P-X-P-S-P-M	PlPSPM
Q01970	PLCB3	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-3	N-P-X-P-S-P	NPtPSP
O43157	PLXB1	Plexin-B1	P-S-P-M-X-S	PSPMaS
P09086	PO2F2	POU domain, class 2, transcription factor 2	P-S-P-X-G-S	PSPqGS
C9JH25	PRRT4	Proline-rich transmembrane protein 4	N-P-X-P-S-P	NPaPSP
Q8WXF1	PSPC1	Paraspeckle component 1	P-S-X-M-G-S	PSqMGS
Q16825	PTN21	Tyrosine-protein phosphatase non-receptor type 21	Q-P-S-P-M-X-S	QPSPMsS
Q7Z5J4	RAI1	Retinoic acid-induced protein 1	N-P-Q-P-X-P	NPQPyP
Q7Z5J4	RAI1	Retinoic acid-induced protein 1	S-P-M-G-S	SPMGS

Q02833	RASF7	Ras association domain-containing protein 7	P-S-P-M-X-S	PSPMaS
Q6WKZ4	RFIP1	Rab11 family-interacting protein 1	N-P-X-P-S-P-M-G	NPvPSMG
Q2KHR2	RFX7	DNA-binding protein RFX7	P-S-P-M-X-S	PSPMsS
O75783	RHBL1	Rhomboid-related protein 1	P-Q-P-S-X-M	PQPSfM
Q7Z6I6	RHG30	Rho GTPase-activating protein 30	Q-P-S-P-X-G	QPSPdG
Q04912	RON	Macrophage-stimulating protein receptor	P-Q-X-S-P-M	PQfSPM
Q8N5L8	RP25L	Ribonuclease P protein subunit p25-like protein	P-X-P-S-P-M	PaPSPM
P24928	RPB1	DNA-directed RNA polymerase II subunit RPB1	P-S-P-M-G	PSPMG
Q5VT52	RPRD2	Regulation of nuclear pre-mRNA domain-containing protein 2	P-S-P-X-G-S	PSPtGS
Q13761	RUNX3	Runt-related transcription factor 3	P-S-P-X-G-S	PSPrGS
Q8N2Y8	RUSC2	Iporin	P-S-P-X-G-S	PSPaGS
Q96LT4	SAMD8	Sphingomyelin synthase-related protein 1	S-P-M-G-S	SPMGS
Q8TEE9	SAP25	Histone deacetylase complex subunit SAP25	P-S-P-X-G-S	PSPcGS
O95487	SC24B	Protein transport protein Sec24B	P-Q-P-S-X-M	PQPSkM
P37088	SCNNA	Amiloride-sensitive sodium channel subunit alpha	P-S-P-X-G-S	PSPgGS
O95810	SDPR	Serum deprivation-response protein	P-S-P-M-X-S	PSPMpS
Q9NP31	SH22A	SH2 domain-containing protein 2A	N-P-Q-X-S-P	NPQySP
O43166	SI1L1	Signal-induced proliferation-associated 1-like protein 1	S-P-M-G-S	SPMGS
O60292	SI1L3	Signal-induced proliferation-associated 1-like protein 3	Q-P-S-X-M-G-S	QPSdMGS
Q15475	SIX1	Homeobox protein SIX1	N-P-X-P-S-P	NPyPSP
Q9NPC8	SIX2	Homeobox protein SIX2	N-P-X-P-S-P	NPyPSP
Q8N196	SIX5	Homeobox protein SIX5	P-Q-P-S-P-X-G	PQPSPqG
Q9UPU9	SMAG1	Protein Smaug homolog 1	Q-P-S-X-M-G	QPSlMG
P51532	SMCA4	Transcription activator BRG1	P-X-P-S-P-M	PpPSPM
P57073	SOX8	Transcription factor SOX-8	P-S-P-X-G-S	PSPaGS
Q9H0E3	SP130	Histone deacetylase complex subunit SAP130	Q-P-X-P-M-G	QPaPMG
Q6ZMY3	SPOC1	SPOC domain-containing protein 1	Q-P-S-P-X-G	QPSPsG
O00267	SPT5H	Transcription elongation factor SPT5	P-X-P-S-P-M	PtPSPM
O00267	SPT5H	Transcription elongation factor SPT5	P-S-P-X-G-S	PSPqGS
Q7KZ85	SPT6H	Transcription elongation factor SPT6	P-X-P-S-P-M	PsPSPM
Q9UQ35	SRRM2	Serine/arginine repetitive matrix protein 2	P-S-P-X-G-S	PSPeGS
Q92502	STAR8	StAR-related lipid transfer protein 8	S-P-M-G-S	SPMGS
Q13586	STIM1	Stromal interaction molecule 1	P-S-X-M-G-S	PSwMGS
Q5T5C0	STXB5	Syntaxin-binding protein 5	N-P-Q-P-X-P	NPQPiP
Q9UMS6	SYNP2	Synaptopodin-2	Q-P-S-X-M-G	QPSkMG

Q9NVA4	T184C	Transmembrane protein 184C	P-S-P-M-G	PSPMG
Q9C0D5	TANC1	Protein TANC1	N-P-X-P-S-P-M	NPpPSPM
Q5QJ74	TBCEL	Tubulin-specific chaperone cofactor E-like protein	P-Q-X-S-P-M	PQgSPM
O60806	TBX19	T-box transcription factor TBX19	N-P-X-P-S-P	NPgPSP
Q6N021	TET2	Methylcytosine dioxygenase TET2	P-S-P-M-X-S	PSPMlS
Q9BXT5	TEX15	Testis-expressed sequence 15 protein	Q-P-S-P-X-G	QPSPsG
P24557	THAS	Thromboxane-A synthase	Q-P-S-P-M	QPSPM
Q495A1	TIGIT	T-cell immunoreceptor with Ig and ITIM domains	P-S-P-X-G-S	PSPpGS
O94900	TOX	Thymocyte selection-associated high mobility group box protein TOX	P-S-P-X-G-S	PSPpGS
Q96NM4	TOX2	TOX high mobility group box family member 2	P-S-P-X-G-S	PSPpGS
Q9NRE2	TSH2	Teashirt homolog 2	P-X-P-M-G-S	PlPMGS
Q9UJ04	TSYL4	Testis-specific Y-encoded-like protein 4	P-Q-X-S-P-M	PQlSPM
Q5TCY1	TTBK1	Tau-tubulin kinase 1	P-S-P-X-G-S	PSPgGS
Q8N6Y0	USBP1	Usher syndrome type-1C protein-binding protein 1	N-P-Q-P-S-P-X-G-S	NPQPSPeGS
O14598	VCY1	Testis-specific basic protein Y 1	Q-P-S-P-X-G	QPSPsG
Q8N398	VW5B2	von Willebrand factor A domain-containing protein 5B2	N-P-Q-P-X-P	NPQPqP
Q9UF83	YM012	Uncharacterized protein DKFZp434B061	S-P-M-G-S	SPMGS
Q6UXV3	YV010	Uncharacterized protein UNQ6126/PRO20091	P-Q-P-S-X-M	PQPStM
Q9Y493	ZAN	Zonadhesin	P-S-P-X-G-S	PSPtGS
Q9Y2Y4	ZBT32	Zinc finger and BTB domain-containing protein 32	P-S-P-X-G-S	PSPyGS
Q86UP3	ZFHX4	Zinc finger homeobox protein 4	P-Q-P-S-X-M-G	PQPSsMG
Q9UPR6	ZFR2	Zinc finger RNA-binding protein 2	Q-P-S-P-X-G-S	QPSPhGS
Q9UPR6	ZFR2	Zinc finger RNA-binding protein 2	P-S-P-X-G-S	PSPhGS
O15231	ZN185	Zinc finger protein 185	P-S-P-X-G-S	PSPaGS
Q9H4Z2	ZN335	Zinc finger protein 335	Q-P-S-P-X-G	QPSPeG
Q96JG9	ZN469	Zinc finger protein 469	N-X-Q-P-S-P	NgQPSP
Q96K83	ZN521	Zinc finger protein 521	P-Q-X-S-P-M	PQvSPM
Q9ULT6	ZNRF3	E3 ubiquitin-protein ligase ZNRF3	P-Q-P-S-P-X-G	PQPSPaG